

# THIS WEEK

## EDITORIALS

**OCEANS** Budget cuts mean that the only way is down for marine science **p.6**

**WORLD VIEW** Carbon offsets are too good to be true **p.7**



**POLLUTION** Noisy gas compressors drive the jays away **p.9**

## Bolstering the link

*Two papers in Nature this week highlight the extent to which human activity is influencing global climate, and underline the need for continued scrutiny of the problem.*

The United States gave Old King Coal a bloody nose last week, with proposals to regulate fossil-fuel emissions that would effectively ban new coal-fired power plants unless they come equipped with technology to capture and store carbon dioxide. The impact in the short term may be minimal: protests against plans for individual new coal power stations and the dash for shale gas as an alternative energy source have forced coal back in the energy pecking order. But, with federal inaction on the climate-change question likely to continue, the move by the Environmental Protection Agency is a welcome one. The agency is finally using the power given to it by the Supreme Court in 2007 to treat carbon dioxide as a pollutant. At the same time, the United Kingdom, which likes to think of itself as an international leader on global warming, seems to be weakening its resolve at home. The Conservative–Liberal Democrat coalition has loosened similar plans to restrict greenhouse-gas emissions from gas-powered electricity generation, although rules that effectively force carbon capture on new British coal power stations remain in place. And, given the government's concomitant failure to announce new mandatory carbon reporting for British business, as required by the 2008 Climate Change Act, it is clear that voters in Britain who were told that they were getting the greenest government ever have been misled.

As the politics of global warming swirls, climate science marches on. This week, two papers in *Nature* underscore what we thought we knew about the problem, as well as offering the kind of surprising result that shows why research must continue apace.

First, as Jeremy Shakun at Harvard University in Cambridge, Massachusetts, and his colleagues show on page 49, carbon dioxide does drive atmospheric warming. Uncontroversial stuff, perhaps, yet the link continues to be questioned by people who would play down the risks of human greenhouse-gas emissions. The queries re-emerged in 2006, when former US vice-president Al Gore showed a graph of historical carbon dioxide levels and temperature in his movie, *An Inconvenient Truth*, and was accused of glossing over the relationship between the two. So let there be no confusion now: the new study confirms that, as Earth emerged from the last ice age some 19,000 to 10,000 years ago, rising global temperatures were preceded by increased global carbon dioxide in the atmosphere — a result that emphasizes the role of carbon dioxide in driving global change in the present day. This relationship is a foundation stone of climate science and of policies to regulate greenhouse-gas emissions, and it is solid.

*Quelle surprise!* the climate sceptics may shout — scientists find proxy data and use a computer model to get the answer they wanted, to seal the conspiracy. Then let the second paper this week show that modern science does anything but offer self-serving results to support existing ideas. For, in a paper published online, Ben Booth and his colleagues at the Met Office Hadley Centre in Exeter, UK, use a different model to question conventional wisdom on how the climate of the North Atlantic Ocean operates (B. B. Booth *et al.* *Nature*

<http://dx.doi.org/10.1038/nature10946>; 2012).

This study looks at the impact of aerosols — such as sulphur dioxide particles ejected by volcanoes and linked to the burning of coal — on sea surface temperatures there. Aerosols that reflect sunlight and can promote brighter cloud formation have been known for some time to affect climate, but the idea has gained appeal in the media during the past decade, under the tag 'global dimming'.

**“Modern science does anything but offer self-serving results to support existing ideas.”**

The study suggests that global dimming could underlie warm and cool periods observed in the North Atlantic basin in the twentieth century — a variability known as the Atlantic Multidecadal Oscillation (AMO). And because the AMO has been implicated in global processes, such as the frequency of Atlantic hurricanes and drought in the Sahel region of Africa in the 1980s, the findings greatly extend the possible reach of human activity on global climate. Moreover, if correct, the study effectively does away with the AMO as it is currently posited, in that the multidecadal oscillation is neither truly oscillatory nor multidecadal.

The implications of this are great, both for study of the climate system and for the impact of policies to control aerosol emissions. It also shows that solid science does not have to be settled science, and that this is no bad thing. ■

## Creative tensions

*Scientists must find ways to improve academic efficiency if they are to keep their independence.*

True scientific creativity is often presumed to be the preserve of independent investigators operating in an environment with none of the practical or political concerns that dog many other workers. In truth, far from being creative, most scientists spend much of their time worrying about funding, sitting in meetings and dealing with administrative bureaucracy.

But, in many ways, academics do live very sheltered lives by today's standards — how many other careers offer tenure or employment for life? And, frequently, an individual's insistence on working as free from tethers as possible can be taken too far, forcing everyone and everything around them to accommodate their needs. This becomes a waste of time and effort — in other words, of money. In this age of economic austerity, has the concept of absolute academic autonomy become a luxury that the scientific enterprise can no longer afford?

A series of Comment articles this week tackles this thorny issue head-on. On page 27, consultant Thomas Marty describes how, in some

departments, faculty members insist on each doing their own course planning, choosing times and subjects independently. This forces the administration to revise courses to ensure that the credits assigned to each are consistent, that students have taken the prerequisites the professors require, and that everything is presented in the format that the computer system recognizes, so that students can register online. Although this may sound like part of administrators' responsibilities, the job can be so big that one department Marty worked with had to dedicate two full-time staff members to resolving such conflicts. Yet academics often grumble at the resources their universities devote to administration.

If scientists truly value their autonomy, they must let go of the traditions that cause more harm than good to the research enterprise. According to Paula Stephan, on page 29, these traditions are often tied to counterproductive financial incentives, such as a US government accounting rule that allows universities to use debt from new construction to increase the indirect rate that they add to grants for overhead costs. This encourages universities to constantly expand rather than house researchers in buildings they already own, and creates an idea that bigger is always better. But, in biomedical sciences, bigger labs have not been associated with a substantial increase in output, and the economic downturn means "the building boom is now costing the scientific enterprise by creating excess space that cannot be paid for," says Stephan.

Scientists may bristle at some of the suggestions proposed to improve the efficiency of the research enterprise. Run academic institutions more like private businesses? Increase the power of institute directors and university presidents so they can make more executive decisions without asking for faculty members' input? Place a 'tax' on the use of temporary workers such as graduate students and postdocs, to encourage scientists to hire more permanent staff scientists?

But scientists should think twice about this instinctive, defensive approach. Something that may seem a threat to academic autonomy is often quite the opposite. A standard template for course planning that all faculty members must adhere to, for example, with strict deadlines for each phase, could cut the number of course revisions. This would free staff to deal with other administrative issues, letting the scientists who had been shouldering that burden get back to research and teaching. Similarly, every 16-person committee that meets once every 2 months

**"Something that may seem a threat to academic autonomy is often quite the opposite."**

for 4 hours can amount to as much as 100 labour days per year, when other costs such as preparation time and staff support are taken into account. Although giving leaders more power to make executive decisions without consulting faculty members may seem to threaten academic independence, in this instance giving up decision-making powers allows scientists to spend more time doing creative, independent research.

And if scientists truly value their power and independence, they must lead the discussion over what works in the research enterprise — and what doesn't. If they don't, someone else will make those decisions for them, by imposing even more funding cuts that directly hurt research and teaching. Scientists cannot continue to live by the double standard that Pierre Azoulay recognizes on page 31, applying deep scepticism to scientific data but unquestioning faith to the practice of science itself. They must approach suggestions to improve academic efficiency with an open mind, trying some and noting whether any impinge on their creativity. However, if implemented properly, none of the suggested changes should have any impact on scientists' all-important academic freedom. If we strengthen the system that supports it, science can only thrive. ■

## Into the depths

*Celebrity missions to the deep ocean won't make up for cuts to marine science.*

Last week, depending on your point of view, film director James Cameron either made history or reached a new low. In descending to the deepest part of the ocean, Cameron became the first human to make a solo visit to the Challenger Deep, nearly 11 kilometres from the surface. He is only the third person ever to make such a dive (see *Nature* <http://doi.org/hsj; 2012>).

Cameron's mission mostly failed in its scientific ambition to recover samples. Equipment failure allowed his submersible to bring back just one piece of sediment and no rocks. But his team has pledged to make further dives, and although there may be no large creatures lurking at that depth, the opportunity to analyse the microorganisms likely to be found in the sediment is unprecedented.

Congratulations flooded in from around the world, many from marine scientists. Among those to send good wishes was the UK National Oceanography Centre (NOC).

But a week that started brightly for the field ended less well. Although Cameron was inspiring the next generation of marine scientists on Monday, by Wednesday news was reaching *Nature* that the NOC was shedding nearly one-quarter of its scientific staff, based at sites in Southampton and Liverpool (see *Nature* <http://doi.org/hsk; 2012>).

This is partly a response to Britain's financial woes, which have kept down or cut budgets at the country's research funding councils in recent years. But the centre's problems have been exacerbated by the stance of the UK Natural Environment Research Council (NERC) — which provides most of the NOC's funding and has decided to "gradually shift the balance of science funding from long-term survey, monitoring and infrastructure such as ships towards front-line, competitively awarded,

strategic environmental research", it said in a statement.

The upshot is that 35 posts are to be lost in the NOC's science section, as the centre attempts to make savings of £3.5 million (US\$5.6 million) a year on its £45-million annual budget.

Ocean researchers around the world have been devastated by the news of the cutbacks, which jeopardize a number of international collaborations in areas such as climate and ocean modelling.

NERC insists that the NOC will still be a major player in the field, albeit "with a somewhat leaner, but even more highly competitive scientific team". But researchers fear that this focus will threaten projects that gestate slowly, which have until now been highly valued.

In the United States, too, researchers have reason to fret. Budget negotiations can be tortuous, with nothing set in stone until the final vote, but there is growing concern about the future of the National Undersea Research Program (NURP). The programme, run by the US National Oceanic and Atmospheric Administration (NOAA), hosts much of the country's research fleet of scientific submersibles. Other NOAA programmes are also under scrutiny.

When the costs of high-profile glamour projects such as manned space flight draw criticism, supporters often say that their public-engagement value offsets mission costs. With Cameron's dive, science got a freebie. No government funds that could have gone to austerity-hit research labs were used: this was one man doing what he wanted with his own money. Others with similar means are set to follow Cameron into the deep, and wealthy individuals are likely to reach space under their own steam and on their own terms in the near future.

But will science be well placed to exploit the massive appeal of Cameron's dive and the new attention that will be given to the ocean depths? Scientists have sent unmanned vehicles to the Challenger Deep between the manned mission in 1960 and last week's visit. Those

trips were made with kit that relies on the skill and dedication of scientists working for programmes such as NURP and the NOC. While Cameron celebrates, ocean science slips a little further out of reach for everyone else. ■

**➤ NATURE.COM**  
To comment online,  
click on Editorials at:  
[go.nature.com/xhunqv](http://go.nature.com/xhunqv)



## The inconvenient truth of carbon offsets

Kevin Anderson explains why he refused to purchase a carbon offset, and why you should steer clear of them too.

**P**lanet Under Pressure was a major conference on the environment held in London last week. As a climate-change scientist, I was invited to organize a session at it and to present my group's research. I declined the offer, and here is why.

The organizers of the conference said that the event would be "as close to carbon neutral as possible". There are good ways to achieve this noble goal: virtual engagement such as video conferencing, advice on lower-carbon travel options, and innovative registration tariffs to reward lower-carbon involvement. But, instead, the organizers chose a series of carbon-offset projects financed through a compulsory £35 (US\$56) fee levied on all delegates.

This was unacceptable to me. Offsetting is worse than doing nothing. It is without scientific legitimacy, is dangerously misleading and almost certainly contributes to a net increase in the absolute rate of global emissions growth.

It is true that the projects funded through these and other offsets can help development. And a rise in emissions from industrializing nations is, in the short term, a good indicator of rising prosperity and should be welcomed.

My objection to offsetting and the Clean Development Mechanism (CDM) — the state-sanctioned version that operates under the Kyoto Protocol — is directed at the claims that they reduce emissions to levels at or below those that would have transpired had the activity being offset not occurred. That spurious argument neglects the various possible impacts of an offset and the repercussions of these for emissions in the longer term.

The science underpinning climate change makes clear that the temperature rise by around the end of this century will relate to the total emissions of long-lived greenhouse gases between 2000 and 2100. Consequently, when considering our impact, we have to look at the total sum of our emissions released in that time; offset projects must be measured over that period. There is no point in reducing emissions by 1 tonne in the short run if the knock-on impact is 2 tonnes emitted in 2020 or even 1.5 tonnes in, say, 2050. The implications of this for the concept of offsetting and the CDM are profound.

For example, if I fly to a climate conference, any claim to offset my emissions must, with a reasonable level of certainty and for a 100-year period, show that the flight emissions plus any emission consequences of the offset projects ultimately sum to zero. It is the immutable impossibility of making such long-term assurances that fundamentally challenges the value of such a claim. Worse still, in a world with rising economic prosperity (fuelled mainly by coal, oil and gas), there is a high probability that offsetting projects contributing to prosperity will increase emissions over and above those arising solely from the activity being offset.

The promise of offsetting triggers a rebound away from meaningful mitigation and towards the development of further high-carbon infrastructures. The UK government's purchase of offsets through the CDM and its simultaneous drive towards both additional airport capacity and the exploitation of UK shale-gas reserves are just two such examples. If offsetting is deemed to have equivalence with mitigation, the incentive to move to lower-carbon technologies, behaviours and practices is reduced accordingly.

Offsetting, on all scales, weakens present-day drivers for change and reduces innovation towards a lower-carbon future. It militates against market signals to improve low-carbon travel and video-conference technologies, while encouraging investment in capital-intensive airports and new aircraft, along with roads, ports and fossil-fuel power stations.

For an offset project to be genuinely low-carbon, it must guarantee that it does not stimulate further emissions over the subsequent century. Although standards and legislation around offsetting and the CDM sometimes consider 'carbon leakage' in the projects' early years, it is impossible to quantify with any meaningful level of certainty over the timeframes that matter. To do so would presume powers of prediction that could have foreseen the Internet and low-cost airlines following from Marconi's 1901 telegraph and the Wright brothers' 1903 maiden flight.

Assume I broke my (self-imposed) seven-year refusal to fly, paid my £35 offset and boarded a plane from Manchester to London for the conference.

In doing so, I add to the already severe congestion at airports, causing delays and allowing politicians to argue for greater airport capacity, arguments only reinforced by the rise in passengers turning to offsets. To meet increasing demand, airlines are encouraged to order new aircraft, which they promise will be more efficient. Feeling pressure, a future government approves new runways, but the extra flights and emissions swamp efficiency gains from the cleaner engines.

Meanwhile, in an Indian village where my offset money has helped to fund a wind turbine, the villagers now have the (low-carbon) electricity to watch television, which provides advertisers of a petrol-fuelled moped with more viewers, and customers. A fuel depot follows, to meet the new demand, and encourages others to invest in old trucks to transport goods between villages. Within 30 years, the village and surroundings have new roads and many more petrol-fuelled mopeds, cars and trucks. Meanwhile, the emissions from my original flight are still having a warming impact, and will do for another 100 years or so.

Where is the offset in that? ■

**CARBON OFFSETTING  
IS WITHOUT  
SCIENTIFIC  
LEGITIMACY  
AND IS DANGEROUSLY  
MISLEADING.**

➔ **NATURE.COM**  
Discuss this article  
online at:  
[go.nature.com/yppmvw](http://go.nature.com/yppmvw)

Kevin Anderson is deputy director of the Tyndall Centre for Climate Change Research at the University of Manchester, UK.  
e-mail: [kevin.anderson@manchester.ac.uk](mailto:kevin.anderson@manchester.ac.uk)



# RESEARCH HIGHLIGHTS

Selections from the  
scientific literature

## NEUROSCIENCE

### High-serotonin mice mimic autism

Some 30% of children with autism spectrum disorder have raised blood levels of the neurotransmitter serotonin, but the relationship between this biomarker and brain function is unclear. Jeremy Veenstra-VanderWeele and Randy Blakely at Vanderbilt University in Nashville, Tennessee, and their colleagues have engineered a mouse with increased levels of blood serotonin that also exhibits behaviours characteristic of autism.

The mouse carries a variant of a gene that encodes SERT, a transporter protein that takes up serotonin into neurons and is also found in blood cells. The authors previously identified this variant in some people with autism. As pups, the mice vocalize less, and as adults, they display repetitive behaviour and are more likely to withdraw when they encounter another mouse. Raised SERT activity may lead to autism-related changes during brain development, the authors suggest.

*Proc. Natl Acad. Sci. USA*  
<http://dx.doi.org/10.1073/pnas.1112345109> (2012)

## NEUROIMAGING

### MRI lights up the whole brain

Functional magnetic resonance imaging (fMRI) can show most of the brain at work during a mental task, if data from large numbers of scans are averaged and different types of analysis are used.

fMRI is typically used by neuroscientists to identify just those few brain areas that are most active during a particular task. However, this approach

might overlook important areas involved in cognitive processing, according to Javier Gonzalez-Castillo at the National Institute of Mental Health in Bethesda, Maryland, and his colleagues.

They imaged the brains of three people who had to decide whether a symbol flashed onto a flickering screen was a number or a letter. Averaging of the results from 100 scans revealed fMRI activations in 95% of the brain that were coincident with the task, varied in shape across regions, and clustered in a highly organized manner. This suggests that the

signals may be functionally relevant rather than noise.

*Proc. Natl Acad. Sci. USA*  
<http://dx.doi.org/10.1073/pnas.1121049109> (2012)

## CARDIOVASCULAR BIOLOGY

### Cell clean-up for artery health

Defects in the intracellular degradation and recycling of cellular components — a process called autophagy — have been linked to several diseases, including cancer. Now two research groups show a link between defective

autophagy and heart disease in mice.

Clay Semenkovich at Washington University in St Louis, Missouri, and his colleagues fed a fat-rich 'Western' diet to mice that are genetically susceptible to atherosclerosis, which is marked by the formation of fatty plaques in arterial walls.

They found dysfunctional autophagy in a type of immune cell called a macrophage, which is abundant in the plaques. Other mice engineered to have autophagy-deficient macrophages and fed a fatty diet had more plaques than



## CLIMATE SCIENCE

### Heads up on a heat wave

A model for forecasting seasonal climate in the United States was able to predict months in advance the intense summer heat that scorched the country's central plains last year (**pictured**).

Lifeng Luo at Michigan State University in East Lansing and Yan Zhang at Scinovation in Princeton, New Jersey, compared the observed number of extremely hot days in the region last summer with temperature-anomaly predictions made by the National Weather Service's upgraded Climate Forecast System, which became operational at the end of March 2011. Almost all the 424 runs of the model between

April and mid-July 2011 accurately predicted an unusually large number of days with extreme temperatures in the region.

As the summer approached, the model became more certain about where, when and how intense the heat wave would be.

Successful prediction of a climate extreme adds confidence to seasonal climate forecasts, the authors say. But how often the model correctly predicts future events still needs to be evaluated, they caution.

*Geophys. Res. Lett.* <http://dx.doi.org/10.1029/2012GL051383> (2012)



those with normal autophagy. Moreover, autophagy dysfunction seems to be associated with activation of inflammasomes — protein complexes that trigger inflammation — in the macrophages.

A team led by Ira Tabas at Columbia University in New York has found that mice that had defective macrophage autophagy and ate a fatty diet showed more signs of advanced atherosclerosis, including plaque necrosis, which has been linked to heart attacks and stroke in humans. *Cell Metab.* <http://dx.doi.org/10.1016/j.cmet.2012.02.011>; <http://dx.doi.org/10.1016/j.cmet.2012.01.022> (2012)

## MATERIALS

## Solvent-free 'ink' glows white

Low-cost, light-emitting inks that can be painted onto surfaces could be a boon for large electronic displays. Researchers have come up with a method of producing glowing ink that avoids the use of solvents, which typically must be evaporated away after the ink is 'printed'. This evaporation can change the ink's colour.

Takashi Nakanishi at the National Institute for Materials Science in Tsukuba, Japan, and his colleagues synthesized small organic molecules that form luminescent liquids at room temperature. These were mixed with other dyes to form a stable paste that glows in shades of white when illuminated with ultraviolet radiation, as illustrated by a light-emitting diode (pictured, left) painted with the paste (right).



The new paste produces an impractical sticky film, but the researchers say they are working to solve the problem. *Angew. Chem. Int. Ed.* 51, 3391–3395 (2012)

## CANCER

## Tumours yield to pressure relief

Lethal pancreatic tumours can be rendered sensitive to chemotherapy by degrading a sugar-based matrix in the tumour that boosts fluid pressure and prevents the inward flow of blood.

Sunil Hingorani at the Fred Hutchinson Cancer Research Center in Seattle, Washington, and his team show that pancreatic tumours in mice have much higher fluid pressures than healthy mouse pancreases owing to high levels of a sugar-based polymer produced by the tumours. An enzyme that chews away the matrix returns fluid pressure to normal, opening up the blood vessels that feed the tumour and allowing chemotherapeutic drugs to penetrate.

Mice treated with a combination of the enzyme and a cancer drug developed fewer tumours and metastases and lived longer — 91.5 days compared with 55.5 — than animals treated with the drug only. *Cancer Cell* 21, 418–429 (2012)

## ECOLOGY

## Noise nixes seed spread

Noise pollution not only alters the behaviour of individual animals, it can also have a wider impact on ecosystems.

Clinton Francis at the National Evolutionary Synthesis Center in Durham, North Carolina, and his colleagues studied how noise levels affect the hummingbird *Archilochus alexandri* and the jay *Aphelocoma californica* (pictured) in an area of New Mexico that has a large

## COMMUNITY CHOICE

The most viewed papers in science

## BIOCHEMISTRY

## Proteins' ticket into the cell

**HIGHLY READ**  
on [pubs.acs.org](http://pubs.acs.org)  
in February

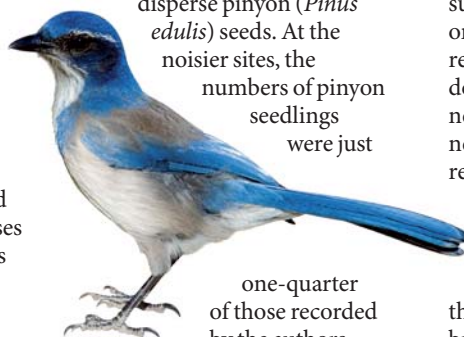
Protein-based drugs are notoriously difficult to deliver into their target cells. Chemists at the University of Wisconsin-Madison have boosted the uptake of a protein by mammalian cells by attaching boronic acid groups to the protein.

Ronald Raines and his team homed in on a modified boronic acid, benzoxaborole, because it has a strong affinity for a sugar molecule abundant on cell surfaces.

The team attached the boronate groups to a protein, RNase A. This enzyme kills the cells that it enters, providing a visible measure of successful entry. Hamster cells internalized the boronated protein four to five times faster than the non-boronated version.

*J. Am. Chem. Soc.* 134, 3631–3634 (2012)

number of gas wells. Increased noise from wells equipped with noisy compressors boosted the pollination of artificial flowers by the hummingbirds, but lowered the number of jays, which disperse pinyon (*Pinus edulis*) seeds. At the noisier sites, the numbers of pinyon seedlings were just



one-quarter of those recorded by the authors at quiet sites. The authors link this shortage to the relative scarcity of jays. *Proc. R. Soc. B* <http://dx.doi.org/10.1098/rspb.2012.0230> (2012)

that makes chilli peppers hot — and its receptor.

Richard Palmiter at the University of Washington in Seattle and his colleagues genetically modified mice so that they expressed a cell-surface receptor for capsaicin only in brain neurons that release the neurotransmitter dopamine, and not in the neurons of the peripheral nervous system that are responsible for the painful response to capsaicin.

When the researchers fed or injected the mice with capsaicin, they found that the compound activated brain dopamine neurons only and stimulated behaviour consistent with increased dopamine release. Similar results were seen in mice with the capsaicin receptor present only on cells that release another neurotransmitter, serotonin.

The researchers could reverse and repeat the effects of capsaicin, as the compound activated neurons for no more than 10 minutes after each administration and did not injure the mice.

*Nat. Commun.* 3, 746 (2012)

## NEUROSCIENCE

## Chilli compound triggers neurons

Neuroscientists have various ways of activating specific neurons in animal brains to determine the cells' roles in neural circuits. A twist on these techniques that its inventors say is less labour-intensive and invasive involves capsaicin — the compound

► **NATURE.COM**

For the latest research published by Nature visit:  
[www.nature.com/latestresearch](http://www.nature.com/latestresearch)

# SEVEN DAYS

The news in brief

## POLICY

### Flu not censored

Two studies that created ferret-transmissible versions of the avian H5N1 flu virus should be published, the US National Science Advisory Board for Biosecurity recommended on 30 March. The board had earlier said that the research should be censored, but its revised stance brings it into line with the World Health Organization. The decision was informed by the release on 29 March of a US government policy that all federal research agencies must review proposals involving specific dangerous pathogens to assess the risk of accidental pathogen-release or deliberate use to damage public health. See [go.nature.com/lgbbjt](http://go.nature.com/lgbbjt) for more.

### Emissions controls

The US Environmental Protection Agency proposed its first standards for greenhouse-gas emissions from power plants on 27 March, finally wielding the authority granted to it by a 2007 Supreme Court ruling that declared carbon dioxide a pollutant under the Clean Air Act. The standards effectively ban new coal-fired power plants that are not equipped to capture and sequester carbon dioxide. They do not cover existing plants, and would allow advanced natural-gas plants to go ahead. See [go.nature.com/rgi2mu](http://go.nature.com/rgi2mu) and page 5 for more.

### Sharing big data

The White House Office of Science and Technology Policy (OSTP) has unveiled a US\$200-million initiative to improve the management, analysis and sharing of the vast data sets that are accumulating as a consequence of federal research. "The world is now generating zettabytes ( $10^{21}$  bytes) of information every

year," said OSTP director John Holdren at a 29 March news briefing. Among efforts announced by various federal agencies, the National Institutes of Health will make data from its 1000 Genomes project, the world's largest collection of human genetic information, publicly available for analysis through the Amazon Web Services cloud computing platform. See [go.nature.com/z2sqlb](http://go.nature.com/z2sqlb) for more.

### No US ban for BPA

The US Food and Drug Administration (FDA) will not ban the controversial chemical bisphenol A (BPA) from food and beverage containers, it announced on 30 March. BPA, a hormone-disrupting plasticizer used in

food-can linings and bottles, has been linked to heart disease, reproductive problems, behavioural problems and cancers. The FDA was forced to rule on the chemical as part of a settlement to a lawsuit from the National Resources Defense Council, an environmental group based in New York. It was not a "final safety determination", the agency emphasized, adding that it still supports research into the safety of BPA. See [go.nature.com/sxxzyn](http://go.nature.com/sxxzyn) for more.

## FUNDING

### Cuts in Spain

Science funding was among areas hit with drastic cuts in Spain's draft budget, which was announced in outline on

30 March. As *Nature* went to press, a full accounting was not available, but central government departments saw their funding slashed by an average of 17%. One government organization, the Centre for the Development of Industrial Technology, which fosters research at technology firms, saw its funding cut by 79%, to €114 million (US\$152 million). See [go.nature.com/yalqgl](http://go.nature.com/yalqgl) for more.

### Ocean job losses

Lay-offs at the United Kingdom's National Oceanography Centre (NOC) could damage international collaborations in the field, oceanographers warned last week. Thirty-five scientific posts — almost one-quarter



TOTAL E&P/UK

## Gas leak in the North Sea

It might take six months to stop a gas leak at the Elgin drilling platform in the North Sea, says the site's owner, oil company Total. The leak, which began on 25 March, is not thought to be environmentally damaging, but demonstrates the uncertainties of offshore drilling. Total says

that the well's main gas reservoir, 5.5 kilometres below the sea bed, was plugged more than a year ago, but gas is emerging from higher pockets, 4 kilometres below the sea bed. The company is considering pumping mud to block the well, or drilling relief wells.

of the science staff — are to be lost at NOC sites in Southampton and Liverpool owing to cuts driven by a squeeze on the centre's main funding agency, the Natural Environment Research Council (NERC). Cuts are also likely at other centres funded by NERC: the British Geological Survey, based in Keyworth, and the British Antarctic Survey, based in Cambridge. See [go.nature.com/lhtgfp](http://go.nature.com/lhtgfp) and page 6 for more.

## Cuts in Canada

Scientists in Canada are dismayed by cuts announced in the country's budget on 29 March. The Conservative government outlined Can\$5.2 billion (US\$5.2 billion) of cuts for the three years up to 2015, and slashed funding for many federal agencies. In particular, it reduced its environment portfolio by 8.3%, including eliminating a federal advisory panel on the environment. The country's main research-granting agencies were told to generate savings to find more money for partnerships between industry and academia, and the National Research Council (which runs federal laboratories) was told to switch its focus to business-driven, industry-relevant research. See [go.nature.com/fqqp5e](http://go.nature.com/fqqp5e) for more.

## RESEARCH

### Biobank opens

The UK Biobank, unusual for its size and depth of data, has opened up its collection to researchers. Scientists can now send in applications to use the data, which include around 10 million samples of blood, urine and saliva collected from more than 500,000 people aged 40–69 years between 2006 and 2010. The biobank has been allocated nearly £100 million (US\$160 million) in funding so far. See [go.nature.com/oshu2i](http://go.nature.com/oshu2i) for more.

### Neutrino finale

The collaboration that last year measured neutrinos travelling faster than light has confirmed two sources of error that explain the findings — and two team leaders have resigned. Analyses posted publicly on 30 March show that a faulty cable caused a 73-nanosecond delay in electrical signals, shortening the neutrino's apparent journey time; a drift in the OPERA experiment's master clock brought the error in arrival time down to 60 nanoseconds. OPERA's physics coordinator, Dario Auterio (pictured), and its spokesman, Antonio Ereditato, have both resigned from the collaboration after months of internal tension



and a vote of no confidence against Auterio by OPERA's board. See [go.nature.com/zkuhip](http://go.nature.com/zkuhip) for more.

### Collider record

The Large Hadron Collider at CERN, Europe's particle-physics lab near Geneva, Switzerland, has collided its first protons at a record 8 teraelectronvolts — the highest energy yet for the machine. The ATLAS detector reported the first test collisions on 30 March. The full experimental run is scheduled to begin around 6 April, and physicists hope to nail down a possible signal for the long-sought Higgs boson by the end of the year.

### Autism increase

Autism prevalence in children aged eight increased by around 80% between 2002 and 2008 in surveyed sites in the United States, according

## COMING UP

### 6 APRIL

The Large Hadron Collider begins its full-power run for the year. <http://atlas.ch/>

### 12–13 APRIL

Researchers and space-policy analysts gather in Strasbourg, France, to discuss whether the International Space Station can be used to test technologies for sending people to Mars. [go.nature.com/q5dool](http://go.nature.com/q5dool)

to a report released by the US Centers for Disease Control and Prevention in Atlanta, Georgia, on 30 March. The report said that by 2008, around 1 in 88 eight-year-old children had been diagnosed with Asperger's syndrome, autism or a related autism spectrum disorder, an increase from 1 in 110 in 2006, and 1 in 150 in 2002. Autism spectrum diagnoses have been increasing for decades, although it is not clear why (see *Nature*, 479, 22–24; 2011).

## PEOPLE

### Physicist in court

After two-and-a-half years behind bars, French-Algerian physicist Adlène Hicheur has appeared in court. In a two-day hearing starting on 29 March, prosecutors argued that Hicheur had plotted terror attacks in conjunction with al-Qaeda's North African branch. Hicheur's lawyers made the case that online exchanges between the researcher and alleged terrorists were not enough to justify his arrest and detention. A decision is expected on 4 May. Prosecutors are seeking a 6-year prison sentence.

► **NATURE.COM**

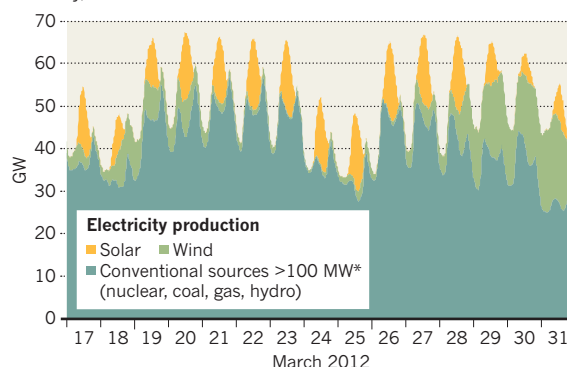
For daily news updates see: [www.nature.com/news](http://www.nature.com/news)

## TREND WATCH

Cuts of up to 29% in subsidies for solar-power producers took effect in Germany on 1 April, in a bid to slow a frantic pace of solar-plant installation. Last year saw an increase of 7.5 gigawatts (GW) in solar capacity, bringing the total to 25 GW; another 3 GW is expected for this year's first quarter. Last year, solar supplied 1–10% of daily electricity demand. Real-time data (see chart) show solar energy partially, but not entirely, smoothing out demand for power from conventional sources by topping up supply during peaks.

### SOLAR PEAKS IN GERMANY

Solar power is helping to smooth out peaks of electricity demand in Germany, but installations will slow now that subsidies have been cut.



\*Small (<100 MW) conventional sources are not captured in real-time data; these typically add 10% to production when accounted for.



# NEWS IN FOCUS

**US POLICY** Effort to protect science from politics hits a bump **p.15**

**SOCIAL SCIENCE** Harvard engineers help to police the mean streets **p.16**

**CLIMATE SCIENCE** Monitoring the vital signs of Asian glaciers **p.19**

**BRAIN IMAGING** fMRI is becoming more than a pretty picture **p.24**



The Jefferson Lab's Free-Electron Laser is a low-cost option in the bid to discover dark-sector forces.

## PARTICLE PHYSICS

# Physicists hunt for dark forces

*Cheap colliders probe debris for hint of 'heavy' photon.*

BY ERIC HAND

In tunnels beneath the Thomas Jefferson National Accelerator Facility in Newport News, Virginia, an accelerator whips a beam of electrons around a racetrack. Their energies are modest, but the beam is tightly packed with them — for it takes a very bright beam to detect a photon that doesn't shine.

In a three-week experiment due to start on 24 April, the electrons will crash into a thin tungsten target at 500 million times a second, creating a cascade of short-lived particles. Amid the debris, physicists with the Heavy Photon Search (HPS) are hoping that they will find signs of something exceedingly rare: a 'heavy' or 'dark' photon. The discovery

would open the door to an unseen world of dark forces and dark atoms that theorists have long speculated about — and could help to pin down the dark matter that is thought to comprise 85% of the matter in the Universe.

The HPS researchers at the Jefferson Lab are quick to concede that the experiment, like two others at the lab probing this dark sector, is a long shot that is likely to achieve little more than null results. But the reasonable price tags for such projects — about US\$3 million to build and run the HPS detector — have prompted more physicists to try. "It's always a great question in physics to go around wondering

if there are more fundamental forces," says physicist John Jaros, co-spokesman for the HPS experiment.

The dark photon, unlike conventional photons, would have mass and would be detectable only indirectly — after the dark photons have decayed into electrons and positrons (the antimatter counterparts of electrons). Yet, like the familiar photon, which carries the electromagnetic force, the dark photon would carry a force — a new fundamental force in addition to the four that we already know about. It would be the first sign of a hidden sector, which could include entire zoos of new particles, including dark matter. "It would be like when Galileo saw moons orbiting Jupiter," says Nima Arkani-Hamed, a theorist at the Institute for Advanced Study in Princeton, New Jersey.

Theorists had hoped that the Large Hadron Collider — the world's highest-energy (and most expensive) particle accelerator at CERN, Europe's high-energy physics lab near Geneva, Switzerland — would open the door to new concepts such as supersymmetry, a set of theories that would resolve some of the problems in the standard model of particle physics. But, so far, it has yielded no clues, such as the dark-matter particles predicted by some supersymmetry models. "The null results are not making people happy," says Philip Schuster, a theorist at Canada's Perimeter Institute for Theoretical Physics in Waterloo, Ontario. "People are wondering what other possibilities are out there."

Instead, some physicists are turning to the 'intensity frontier' — creating many collisions and teasing rare events from the wreckage. The electron beams at the Jefferson Lab are not the most powerful, but they are extremely intense.

The idea for a dark sector was first proposed in 1986 (B. Holdom *Phys. Lett. B* **166**, 196–198; 1986), but remained largely unexplored until a group of theorists, including Arkani-Hamed, resurrected the theory a few years ago (N. Arkani-Hamed *et al. Phys. Rev. D* **79**, 015014; 2009). The group embellished the idea in light of results from a 2006 satellite mission called PAMELA (Payload for Antimatter Matter Exploration and Light-nuclei Astrophysics), which had observed a puzzling excess of positrons in space. Theorists suggested that they might be spawned by dark-matter particles annihilating each other. But the heavy particles most often suggested (WIMPs, weakly interacting massive particles) would have also decayed into ►

► **NATURE.COM**  
For more on direct  
detection of dark  
matter:  
[go.nature.com/gxqfyf](http://go.nature.com/gxqfyf)

► protons and antiprotons, which weren't seen by PAMELA. A dark-matter particle from the dark sector — “even darker matter”, quips Arkani-Hamed — would be seen only through a decay involving the force-carrying dark photon, which would make positrons but not antiprotons.

Another motivation came from an intriguing result reported in 2004 by physicists at Brookhaven National Laboratory in Upton, New York. They found that the magnetic moment created by the spin and charge of the muon, a short-lived particle similar to an electron, did not match the predictions of the standard model. This anomaly, called the muon g-2, could also be rectified by a dark-sector force, says Arkani-Hamed. He adds that the idea is not as crazy as it sounds. “The whole set-up is totally vanilla and conservative from a theorist's point of view,” he says.

The predictions can be tested cheaply and relatively quickly. The main 6-giga-electronvolt electron beam at Jefferson Lab has the right energy to probe the most likely mass range for a heavy photon. After the HPS's three-week test run, the beam will be shut down for an upgrade that will double its energy. This will allow the HPS and another project, the A Prime Experiment (APEX), to explore other parts of the dark sector in 2015. A third proposal, called DarkLight, would use the beam that drives the lab's free-electron laser to look for heavy photons at lower energies (see ‘Feeling in the dark’).

Arkani-Hamed says that he won't be surprised if the future path of particle physics emerges from modest experiments such as those at the Jefferson Lab, rather than from work at CERN. “It could be that these much smaller, faster, cheaper, upstart, high-intensity, low-energy experiments might actually dig up evidence for new physics before the big monsters.” ■

## FUNDING

# Cyprus Institute loses money and support

*Budget withheld as audit raises concerns.*

BY ALISON ABBOTT

Once billed as the MIT of the Middle East, the Cyprus Institute in Nicosia is now so deep in crisis that it has not been able to pay last month's salaries. The Cypriot parliament is withholding the institute's 2012 budget amid a storm of criticism — from an audit exposing poor financial management to complaints about the management style of the institute's president, nuclear physicist Costas Papanicolas.

Last week, parliamentarians called for an independent analysis of the institute's activities to help to decide whether to continue financial support. The institute employs nearly 100 staff, but others have left for more secure environments. “If nothing constructive happens soon, the institute will collapse,” says atmospheric chemist Jos Lelieveld of the Max Planck Institute for Chemistry in Mainz, Germany, who holds a joint position at the Cyprus Institute.

The multidisciplinary institute was conceived eight years ago by Hubert Curien, a former president of the CERN council, and John Joannides, then chief executive of the Cyprus Development Bank. Their aim was to create a high-quality research university that could bridge science communities in Europe and the Middle East, and build local research capacity. To sidestep public-sector bureaucracy, it was organized as a private, non-profit foundation.

The government provided seed money in 2005. The plan was for the institute to have six research centres and 2,500 graduate and undergraduate students by the end of 2009, and for it to be self-financing through fundraising, grants and student fees by 2010.

Three research centres — in energy and environment, high-performance computing and archaeology — were launched in 2007. The institute formed ties with international institutions such as the Massachusetts Institute of Technology (MIT) in Cambridge and the University of Illinois at Urbana Champaign to help to plan research and offer training programmes. It also signed an agreement worth more than €3 million (US\$4 million) with MIT to design a solar-energy pilot plant able to generate electricity and desalinate seawater.

But in 2010, politicians, aware that the vice-presidents for research and for operations had resigned, began to question why the institute was still receiving public money. The institute

had not expanded the number of research centres and the proposed student population had not emerged to provide the anticipated US\$62.5 million annual income. Moreover, no private donors had come forward.

“Naturally, we were concerned that an institute intended to be self-supporting was absorbing around €8 million of public money a year without the checks and balances that a public organization would have,” says parliamentarian Nicholas Papadopoulos, one of the institute's most aggressive critics. Cyprus invests too little in research, he says, but “its small investments still need to be efficient”.

In March 2011, parliament called in the auditor general. The audit report, delivered on 17 February this year, describes the business plan of the Cyprus Institute as unrealistic and over-ambitious, and criticizes a lack of transparency in its financial accounting, and in the hiring and paying of staff. It also notes that one-fifth of the staff left during 2011.

A number of scientists associated with the institute told *Nature* that they found Papanicolas' micro-management style intimidating.

**“If nothing constructive happens soon, the institute will collapse.”**

They say it hindered their work and is holding the institute back, although most asked not to be identified. Andrea Pozzer, who left last year for a

post at the International Centre for Theoretical Physics in Trieste, Italy, says “the interdisciplinary approach of the institute was exciting. But the general atmosphere was often unhappy and there were constant rumours about poor management that were not motivating”.

Papanicolas denies that he interfered in day-to-day scientific activities.

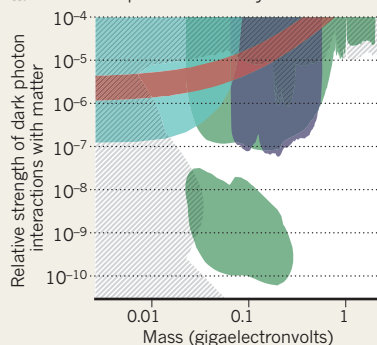
“I agree it may not be easy to work with Costas Papanicolas if you don't gain his trust or live up to his high expectation,” says theoretical physicist Edouard Brézin at the École Normale Supérieure in Paris, and chairman of the Cyprus Institute's board of trustees. But he adds that the institute should be more transparent.

The parliament's public-accounting committee endorsed the auditor's report last week and asked the finance committee to release the first quarter of the institute's 2012 budget. But Papadopoulos, who chairs the finance committee, is postponing discussions about this. ■

## FEELING IN THE DARK

Three experiments will search unexplored mass regions for a dark photon, which could explain why muons flout the standard model.

**Experiments:** ■ DarkLight ■ APEX ■ HPS  
■ Where muon data hint dark photon may be  
■ Where dark photon is already ruled out







Salmon deaths in the Klamath River have prompted legislation to remove dams and restore water flow.

## POLICY

# US integrity effort hits troubled water

*Allegations by integrity officer who lost his job are a setback for plan to quash political inference in science.*

BY EUGENIE SAMUEL REICH

When US President Barack Obama announced a government-wide effort to protect federal science from political interference, the US Department of the Interior (DOI) took an early lead. In 2011, it became the first agency to finalize a new policy on scientific integrity and it has hired ten scientific-integrity officers to work with staff in its various bureaus. But the DOI may also be the first to run into a problem with the way the policies are implemented, as one of those officers claims to have been fired for upholding the guidelines.

"I thought I was doing the job I was hired to do and was doing the right thing. I was stifled," says Paul Houser, a hydrologist at George Mason University in Fairfax, Virginia, who was appointed as scientific-integrity officer for the DOI's Bureau of Reclamation in April 2011. Houser was fired on 10 February and filed a complaint under the DOI's

scientific-integrity policy two weeks later.

Houser alleges that he lost his job because he raised concerns with the DOI about the way it had represented the science behind a plan to remove four hydroelectric dams from a stretch of the Klamath River that straddles the California–Oregon border. DOI spokeswoman Kate Kelly declined to say why Houser had been fired. But watchdog groups say that Houser's complaint underscores the need for protections for those tasked with enforcing the new policies.

"There are things the president has to lead on and whistle-blower protection is one of them," says Francesca Grifo, director of the Union of Concerned Scientists' scientific-integrity programme in Washington DC.

Houser says that his concerns conflicted with the DOI's desire to build a strong public case for removing the dams. Environmental groups support the removal because the dams reduce water flow in the river and have been blamed for large numbers of salmon deaths.

Secretary of the Interior Kenneth Salazar was planning to decide whether the dams should be removed by 31 March, but legislation authorizing removal has stalled in the US Congress.

Houser says that he was asked by a press officer to check some material that the DOI planned to make public about the probable environmental impact of the dams' removal. But the material painted an overly rosy picture of the benefit, Houser says. For example, in a summary document, the DOI said that studies had shown that the annual production of Chinook salmon (*Oncorhynchus tshawytscha*) would rise by 83% a year after the dams were removed. However, it did not include any of the uncertainties about how the population would respond that an expert panel commissioned by the DOI had listed. In the final version of the summary — which is now on a government website — the number was changed to 81.4%. "That number expresses an accuracy that's ludicrous," says Houser. The figure comes from an unpublished computer-modelling study and had an uncertainty range of –59.9% to 881.4%, which was not reported in the summary.

Houser says that last September, his supervisor, deputy commissioner for external and intergovernmental affairs Kira Finkler, chided him for documenting his concerns. He says Finkler told him that "the secretary wants to remove those dams". Finkler did not respond to questions from *Nature* about the situation, but the scientific-integrity officer who is overseeing implementation of the department's policy, Ralph Morgenweck, confirms that Houser's complaint is being investigated.

Kelly says that the DOI is looking forward to the outcome of the investigation. "We believe all actions will be proven to be fully justified," she says, adding that the studies the agency is using about the impact of the dam removal are available on the Internet for anyone to see and review. However, members of the expert panel contacted by *Nature* have said that they, too, felt that the materials flagged by Houser played down the uncertainties in scientific predictions.

The 30 March deadline for US science-related agencies to finalize their integrity policies has now passed, but cases such as Houser's could set the tone for how well the effort takes root, says Grifo. "These early examples are incredibly important."

Both Grifo and Jeff Ruch, executive director of the watchdog group Public Employees for Environmental Responsibility in Washington DC, believe that an absence of whistle-blower protections is likely to make the policies less effective. And for Houser, the departmental hierarchy may not have helped: the scientific-integrity officers report to the heads of the bureaus they work with, which seems to leave the officers vulnerable and could weaken their ability to uphold policies. Morgenweck says that the DOI may reconsider the reporting line when it next revises its policy. ■





SGT SEAN MURPHY, MA STATE POLICE

In Springfield, Massachusetts, police draw on military counter-insurgency methods to combat high rates of crime and gang-related violence.

## SOCIAL SCIENCE

# A data-driven war on crime

*Scientific tools inform a unique combination of military tactics and police work.*

BY SHARON WEINBERGER

A routine midday police patrol through the crime-ridden North End of Springfield, Massachusetts, is interrupted when state troopers spot a teenager riding an unregistered motorcycle with no helmet. Instead of arresting the underage rider, who initially tries to flee, the police speak to him and his father, who is in the crowd that has gathered at the scene. They agree not to charge the boy, for now, if he will tell them who gave him the motorcycle and write a two-page paper about obeying the law.

This approach to law enforcement is part of an initiative called Counter Criminal Continuum (C<sup>3</sup>), which draws on strategies used in US military campaigns against insurgents in Iraq and Afghanistan. The Massachusetts state police launched C<sup>3</sup> in 2009 as part of an effort to combat the drugs-and-gang-related violence that run rampant in parts of Springfield, one of the most dangerous cities in the United States. Now, undergraduate students at Harvard University in Cambridge, Massachusetts, are attempting to bolster the initiative with state-of-the-art data-gathering and analysis of how well these counter-insurgency techniques are working.

Led by Kevin Kit Parker, a biomedical

engineer and professor at Harvard, engineering students are collecting data on encounters such as the one with the boy on the motorcycle to look for correlations between policing tactics and crime rates.

Parker, who advises the Pentagon on research projects, has pushed the US defence department to adopt scientific ways of evaluating the effectiveness of counter-insurgency (COIN) in conflict-torn areas. Collecting precise data on factors such as murder rates or income levels in a war zone is almost impossible, however. "There was no way to do science and technology to back up COIN," Parker says. But Springfield, where the police were adopting COIN techniques, offered a real-world laboratory in which to measure their effectiveness.

Parker, a US Army major who had recently returned from Afghanistan, learned of the programme shortly after it began from Michael Cutone, a Massachusetts state trooper and US Green Beret who served in the same National Guard unit as Parker. When Cutone returned from Iraq, he realized that the chaos he saw in Springfield bore a strong resemblance to what he had seen in the northern Iraqi city where he had been deployed.

In 2009, Cutone approached John Barbieri, Springfield's deputy chief of police, with the

idea of taking the US counter-insurgency doctrine — which relies on tactics such as collecting intelligence to take down insurgent networks and winning support from the local population — and applying it in Springfield. "I think we can use the Army Special Forces counter-insurgency model and adapt it for law enforcement to combat and attack the gangs and drug dealers," he told Barbieri.

Barbieri decided that it was worth a shot. "If you've got something that works in a remote village," he recalls telling Cutone, "then it's got to work here."

When Parker learned of the effort, he seized on it as a way to test some of the theories that the military was already spending billions of dollars to apply. The Springfield Police Department agreed to work with Parker and his team, and students in Parker's class have routinely been making the two-hour trip to Springfield to gather data on everything from arrests and shootings to education level and rates of infection with sexually transmitted diseases.

Among the elements that the Springfield C<sup>3</sup> programme borrows from the military is a data-collection system called Lighthouse, which researchers at the US Naval Postgraduate School in Monterey, California, developed to gather data on insurgents in countries such as Afghanistan. The Springfield police are

using Lighthouse to collect information on gang members and criminals. Parker's students are trying to use this and other crime-related data to help police to improve their methods. They have created a 'war room', where they are working on social-network analysis and computational social science in an effort to predict where crime may happen.

The students use a variety of methods, including organizing street-level data with Google Maps to plot and photo-document neighbourhoods block by block, and then combining the data with public statistics for crime, demographics and public health. They then use linear-regression techniques to analyse the statistics with the aim of developing computational tools to help police anticipate crime or identify key gang leaders.

In some respects, the project embodies former US defence secretary Robert Gates's vision of strengthening the ties between the Pentagon and academic research. But the ultimate question for Springfield — as for Afghanistan — is whether counter-insurgency (or counter-gang) strategies are effective. Barbieri points to the results in Springfield: a 62% drop in crime over the first year that C<sup>3</sup> was implemented. "I am stunned by how well it works," he says.

On the ground, success is not always obvious. The optimism of the scene in which the

teenager on the motorcycle shakes hands and smiles with state troopers dissolves minutes later, when a police dispatcher reports that the motorcycle was stolen. The teenager now faces possible criminal charges for theft unless he reveals who gave him the bike (which he ultimately does).

And even if C<sup>3</sup> works, not everyone believes it should be called counter-insurgency. John Sullivan, a lieutenant with the Los Angeles Sheriff's Department, says that many of these tactics originated with the police, and he worries that calling the approach counter-insurgency gives the false impression that the police are using military methods. "That convolutes the discussions," he says.

Hugh Gusterson, an anthropologist at George Mason University in Fairfax, Virginia, and a frequent critic of the military's involvement in social-science research, warns that the idea of applying a counter-insurgency approach in domestic law enforcement, however it is labelled, risks casting local communities as hostile populations.

"I'd feel that universities were making more of a positive contribution to society if they

**Students are developing computational models to analyse and improve police methods.**

worked instead on understanding the connections between crime, social inequality and perceived injustice," he says.

Whether it is police doctrine or military doctrine, studying the effectiveness of counter-insurgency techniques domestically makes sense, says Douglas Ollivant, a senior fellow at the New America Foundation based in Washington DC. Ollivant, who worked as a counter-insurgency adviser to the US military, says that surveys can be carried out and data collected more easily in the United States than in war zones, and could help to answer questions for both the military and the police. "Has COIN been effective?" asks Ollivant. "I think there's a dearth of evidence."

Yet the military has been reluctant to fund efforts like the Springfield-Harvard collaboration. As an adviser to defence agencies, Parker has pitched the idea of a domestic science-and-technology lab, so far without success. Although what emerges from his work on the Springfield programme may strengthen his case, he acknowledges that a domestic theory of counter-insurgency is a tough sell when many believe it has failed abroad.

"I feel like I'm still trying to find some part of this war I can win," he says, referring to Afghanistan. "Maybe the only thing I can win is building the database and tools for the next insurgency we have to fight." ■



The fate of many glaciers in Tibet and its surrounding mountain ranges remains controversial.

## CLIMATE CHANGE

# Glaciologists to target third pole

*High-tech stations on track to monitor third-largest ice store.*

BY JANE QIU

With the health of the world's highest glaciers in dispute, an international team is planning a long-term campaign to measure the vital signs of the ice atop Tibet and its surrounding mountains.

The 46,000 or so glaciers in the region — known as the Third Pole — supply water to some 1.4 billion people in southern and central Asia, and although many climate studies suggest that the ice is disappearing fast, not all measurements are so dire. Now, the Third Pole Environment (TPE) programme, an international effort to assess the effect of climate change in the region, aims to get some answers by monitoring 25 of the glaciers, which project leaders plan to identify in coming weeks.

Starting later this year, a team led by researchers from Asia will survey the glaciers twice a year and use satellite measurements to look for changes in mass balance — the sum of the snowfall that builds up the glaciers and the melting that shrinks them.

The study sites have been chosen to help tease apart the key factors in a glacier's fate — such as elevation, topography,

geographical setting, climate and the type of debris that covers the ice. “The flagship glaciers are pieces of the puzzle of climate responses on the Third Pole,” says Yao Tandong, director of the Chinese Academy of Sciences' Institute of Tibetan Plateau Research (ITP) in Beijing, and chairman of the TPE's science committee. Using standardized methods at each site, the team hopes to discern how glaciers in different parts of the Third Pole are responding, and what is driving the changes. The team will put the information collected during the study into a public database after a proprietary period of a couple of years, says Yao.

A 2007 report from the Intergovernmental Panel on Climate Change projected that Himalayan glaciers could disappear as early as 2035 — but that claim turned out to be baseless. “The exact health status of the glaciers is still an unsettled issue,” says Shresth Tayal, a glaciologist at the Energy and Resources

Institute in New Delhi. Researchers currently tend to rely on satellite measurements of the region's glaciers to keep

tabs on the glaciers' surface areas and end points. “This can be misleading,” says Tian Lide, a glaciologist at the ITP, who has been conducting field measurements of glacier mass balance for two decades. “Some glaciers may have the same or even increased surface area but are in fact thinning.”

A 2010 study using measurements taken by the Gravity Recovery and Climate Experiment (GRACE) satellite mission indicated that the Third Pole is shedding roughly 50 gigatonnes of ice per year<sup>1</sup>. And an unpublished inventory of Tibetan glaciers led by Liu Shiyin at the Cold and Arid Regions Environmental and Engineering Research Institute in Lanzhou, China, shows that more than 70% of the glaciers on the Tibetan plateau are retreating. But an analysis<sup>2</sup> this year of GRACE data suggests that overall, high-altitude Asian glaciers are losing ice only one-tenth as fast as the previous estimates, and that those in the Tibetan Plateau are actually growing on average.

## HIGH-STAKES DATA

Yet many glaciologists are sceptical about the latest GRACE results. “When satellite data are in stark contrast to what many glaciologists have experienced through decades of field research, one must question their validity,” says Pradeep Mool, a remote-sensing expert at the International Centre for Integrated Mountain Development in Kathmandu, Nepal.

Detailed analyses of 25 glaciers will not settle the controversy, but “they're a good starting point”, says glaciologist Koji Fujita at Nagoya University in Japan. As well as assessing mass balance, the team will set up several comprehensive observatories to monitor the weather and solar radiation and measure properties of the snow, soil and ice, says Daqing Yang, a hydrologist at Environment Canada in Gatineau, who is involved in the study. It will also test methods for measuring snow amounts at high elevations — “a missing but badly needed piece of information in mountain research”, says Yang.

Imtiaz Rangwala, a climatologist at Rutgers University in New Brunswick, New Jersey, who is not involved in the TPE, says that the stations may help to resolve a pressing problem about climate change in high regions. Many climate simulations suggest that higher elevations will warm faster than lower ones, but Rangwala and his colleague James Miller reported last month that many mountain regions do not follow such a clear pattern<sup>3</sup>. The campaign, he says “will bridge a major knowledge gap in mountain research, especially at a time when high-elevation observatory stations elsewhere are at risk of being closed down due to lack of funding” ■

1. Matsuo, K. & Heki, K. *Earth Planet. Sci. Lett.* **290**, 30–36 (2010).
2. Jacob, T., Wahr, J., Pfeffer, W. T. & Swenson, S. *Nature* **482**, 514–518 (2012).
3. Rangwala, I. & Miller, J. *Clim. Change* <http://dx.doi.org/10.1007/s10584-012-0419-3> (2012).

➔ **NATURE.COM**  
For more on the third pole, see:  
[go.nature.com/gwql2z](http://go.nature.com/gwql2z)



## EDUCATION

# PhDs leave the ivory tower

UK doctoral training centres prepare students to run a lab — or work outside academia.

BY DANIEL CRESSEY

Not so long ago, doctoral students were viewed as the galley slaves of the scientific world, spending long hours in the lab for a meagre wage and the promise that three precious letters — PhD — would eventually burnish their name.

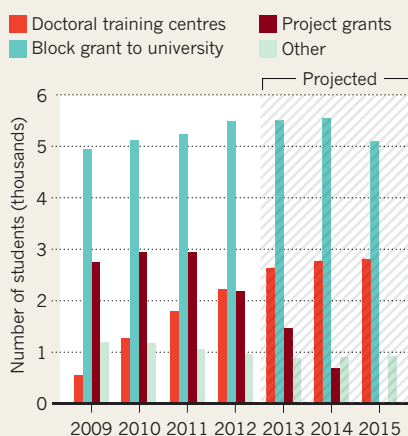
But that attitude has changed. Recognizing that few graduates spend their entire careers at the bench, research funders and education authorities are reshaping the PhD to train students in non-science skills such as networking as well as research. One of the most radical expositions of this philosophy is unfolding in the United Kingdom, where PhD students are increasingly coming out from under the wing — and the shadow — of a PhD supervisor. Instead of being trained individually in one academic's research group, they are being taught in cohorts in a doctoral training centre (DTC) — a university-based hub focusing on highly specific areas, such as chemical synthesis or nuclear fission. DTC courses last four years rather than the three of a standard UK PhD, and include formal coursework as well as lab experience.

Britain's main physical-sciences funding body, the Engineering and Physical Sciences Research Council (EPSRC), has opened more than 50 DTCs, most of them in 2009. This year, for the first time, it has funded more students through centres than through research grants (see 'Training in transition'). Some academics worry that the centre-based approach will squeeze blue-sky research out of PhD programmes — and deprive senior scientists of eager assistants. But delegates gave positive reviews of the DTC model last month at a meeting in London

***"The world of research has changed, is changing and will continue to change. Training centres are a way to respond to that."***

## TRAINING IN TRANSITION

At the UK Engineering and Physical Sciences Research Council, doctoral training centres look set to take over from project grants as a way to fund PhD students.



organized by Vitae, a government-funded group based in Cambridge that pushes for improved researcher training.

"The world of research is changing, has changed and will continue to change. I see centres as being a way that we respond to that," says Neil Viner, associate director of the EPSRC. Other British funding agencies are now following the EPSRC's lead, and several funders abroad are either considering adopting similar systems, or already working on it.

Rudiger Woscholski, deputy director of Imperial College London's Institute of Chemical Biology DTC, says that the centre's students are far better equipped than traditional PhD students to collaborate with other labs and with industry. Arie Iserles, co-director of the Cambridge Centre for Analysis at the University of Cambridge, adds that the four-year DTC programmes also leave graduates better prepared for postdoctoral work.

Similar schemes have been launched by

other UK research councils, such as the Biotechnology and Biological Sciences Research Council (BBSRC), which sponsors doctoral training partnerships in which partner institutions receive grants to coach cohorts of students for four years. Although the partnerships do not create centres as such, they are designed to build a structured doctoral training environment. Celia Caulcott, director of innovation and skills at the BBSRC, points out that the councils are also keeping other funding routes open. Students can still pursue focused research-based degrees, she says, and universities that fail to become partners can still maintain PhD programmes funded by block grants based on their research income.

Other nations are closely watching Britain's PhD reform. Lidia Borrell-Damian, senior programme manager for research and innovation at the European University Association in Brussels, says that there has been a Europe-wide move towards more formalized PhD training and stronger links with industry. Germany, for example, has begun to shift its PhD programmes to a model that increases institutional oversight and aims to prepare students for the wider world of work (see *Nature* **472**, 276–279; 2011). The United Kingdom, adds Borrell-Damian, is a leader in many other innovative aspects of PhD reform, such as co-supervision with an academic and an external supervisor.

"In the years to come almost all universities in Europe will have to some extent a graduate school or a doctoral school," she says. "That is unstoppable." ■

## CORRECTION

The News story 'Chemistry's web of data expands' (*Nature* **483**, 524; 2012) wrongly stated that Antony Williams led the effort to clean up structures on Wikipedia. Although he took part in the effort, he did not lead it.

CHRISTIE'S IMAGES/CORBIS



## TOP STORY

Gene behind mutations in van Gogh's sunflowers pinpointed  
[go.nature.com/iecrsb](http://go.nature.com/iecrsb)



## OTHER NEWS

- Moving towards a universal flu vaccine [go.nature.com/vlg8eb](http://go.nature.com/vlg8eb)
- Humans have been cooking with fire for one million years [go.nature.com/cfwnav](http://go.nature.com/cfwnav)
- Ratio rethink will adjust rock clocks [go.nature.com/nwbdao](http://go.nature.com/nwbdao)

## PODCAST



A giant feathered dinosaur, financial incentives for research and a disease that crosses oceans  
[go.nature.com/va1wve](http://go.nature.com/va1wve)



# *BLOWING IN THE WIND*

*The mysterious Kawasaki disease might cross the Pacific on air currents high in the atmosphere.*

BY JENNIFER FRAZER

**T**he desperately ill baby had been airlifted in from Wyoming, recalls Jane Burns, thinking back to 1981 and her third year as a paediatric resident at the University of Colorado School of Medicine in Denver. Twenty-one days later, the little girl's skin rashes were mostly gone, but the accompanying fever was still raging, and Burns had no idea why.

"I think this is Kawasaki disease," said Richard Anderson, an infectious-disease fellow at the school, who had also examined the tiny patient.

Burns was stunned. Kawasaki disease was uncommon even in Japan, where it had been first identified in the early 1960s, and was almost unheard of in the United States. It was also utterly mysterious — some kind of inflammation in the blood vessels that primarily

targeted children under the age of five and produced a variety of dramatic symptoms (see 'Mysterious malady'). Burns had heard of the disease only because she had encountered two Kawasaki disease patients in the previous year. And now Anderson was telling her she'd just got a third.

Less than 12 hours later, the baby was dead.

"I was so completely amazed by this disease that had now come and got me for a third time that I did the post mortem with the pathologist," says Burns. "And I'll never forget opening up her chest and looking at her heart and seeing the aneurysms sitting there."

These balloon-like bulges are common enough in the blood vessels of adults, especially those with risk factors such as diabetes or high blood pressure. But in a once-healthy

infant? For Burns, now a paediatrics researcher at the University of California, San Diego (UCSD), and director of the Kawasaki Research Center at the UCSD Children's Hospital, the baby's death was a turning point. She has been studying Kawasaki disease ever since, whenever she can find funding. And she is not alone. "Kawasaki disease is something that has been fascinating people in infectious disease since it was described," says Ian Lipkin, an epidemiologist and director of the Center for Infection and Immunity at Columbia University in New York. "It smells like an infectious disease; we've just never been able to catch the culprit."

Epidemiologists now have a new place to look: on winds blowing from central Asia. A team of medical and climate scientists, including Burns, argue in last November's issue of

**Seasonal winds from central Asia could be bringing Kawasaki disease into Japan.**

C. KOBER/ROBERT HARDING/CORBIS

**NATURE.COM**  
Read more about  
microbes in the  
atmosphere at:  
[go.nature.com/5t8d3n](http://go.nature.com/5t8d3n)

*Scientific Reports*<sup>1</sup> that the agent of Kawasaki disease is not only reaching Japan from the Asian mainland by this route, but it seems to be crossing the Pacific Ocean to infect children in Hawaii and the North American mainland.

If windborne spread turns out to be true, the Kawasaki disease agent will be the first viable human disease pathogen proved to cross thousands of kilometres of ocean by natural means (as opposed to carriage on planes or ships). And it may not be the last: researchers are beginning to ask whether the wind might also be a factor in the spread of influenza.

## FIRST SIGNS

Japanese paediatrician Tomisaku Kawasaki saw his first case of the disease in 1960. He had no idea what it was. But it was so striking that he made diagrams and kept detailed records of the symptoms.

"He had this joke that he filed it in a folder called GOK, which was God Only Knows," says Burns.

Kawasaki published the first formal description of the malady in 1967 in Japanese<sup>2</sup>. Since then, there have been three major outbreaks in Japan: peaking in April 1979, May 1982 and March 1986. And the number of cases has been rising steadily each year (see 'Seasonal cycle'), despite Japan's falling birthrate. Today, the average annual incidence of about 12,000 cases in Japan rivals that of the earlier outbreaks at their peaks. Even in San Diego, Burns sees 80 to 100 new cases a year.

It is hard to say how widespread the disease actually is, because outside Japan the diagnosis is probably missed as often as not; children get mysterious rashes all the time. But the consequences can be serious. The kind of aneurysms that Burns saw in the 1981 autopsy occur in a quarter of untreated cases, and kill about 1 patient in 100. In addition, the body's attempts at repair can cause a build-up of scar tissue or narrowing of the arteries.

"Damage can be completely silent for decades until it's not," Burns says. "And the presentation can be with a massive heart attack in these young adults, who just have no idea why this is happening to them."

Genetics seems to play a big part, she says. Asian children are more susceptible than those

in other ethnic groups. The immune system is deeply involved as well. The disease is characterized by widespread inflammation, which eventually focuses on the smooth muscle cells found in the walls of medium-sized arteries, in the heart and elsewhere, and can lead to aneurysms. The only effective treatment is intravenous injection of human immunoglobulin G, the major antibody fraction in blood, which — for reasons that are still not completely understood — lowers the probability of aneurysms from 25% to 1–5%.

Yet the cause of the disease remains GOK: for decades, researchers have looked for the Kawasaki disease agent among viruses, bacteria and every other category of pathogen, to no avail. Lipkin has been collaborating with Burns on this search for 20 years, applying each new molecular diagnostic tool as it came along and coming up empty-handed each time. Burns herself spent much of the

between 1987 and 2000. A trend was clear: cases were sharply seasonal, peaking in the winter and early spring, and again in early summer, which suggested that an environmental factor was involved<sup>3</sup>.

Funding ran out and that lead went cold. But then in 2007, Cayan heard a lecture by Xavier Rodó, director of the Catalan Institute of Climate Sciences in Barcelona, Spain, who was on sabbatical at UCSD. Rodó had experience in figuring out how climate affects infectious diseases such as cholera, and had designed mathematical and statistical tools to pick out variables that might have low signal-to-noise ratios, or that might be intense but brief.

After the lecture, Rodó recalls, Cayan told him about Burns and her access to the extraordinary database of Japanese Kawasaki disease patients. "He said many people have had a go at this disease but no one really found anything that significant," says Rodó — a challenge too good to refuse.

Adding Japanese records of more than 247,000 patients to his programs, Rodó let the software plough through a plethora of climate variables, including temperature and humidity. One trend popped out: when the winds blew from central Asia across Japan, the number of Kawasaki disease cases skyrocketed. All three major outbreaks in Japan had followed this pattern, and it was also evident in the normal disease seasons. When the winds shifted to blow from the Pacific, cases dropped<sup>1</sup>. And when winds from central Asia made their way to Hawaii or California, cases spiked there too.

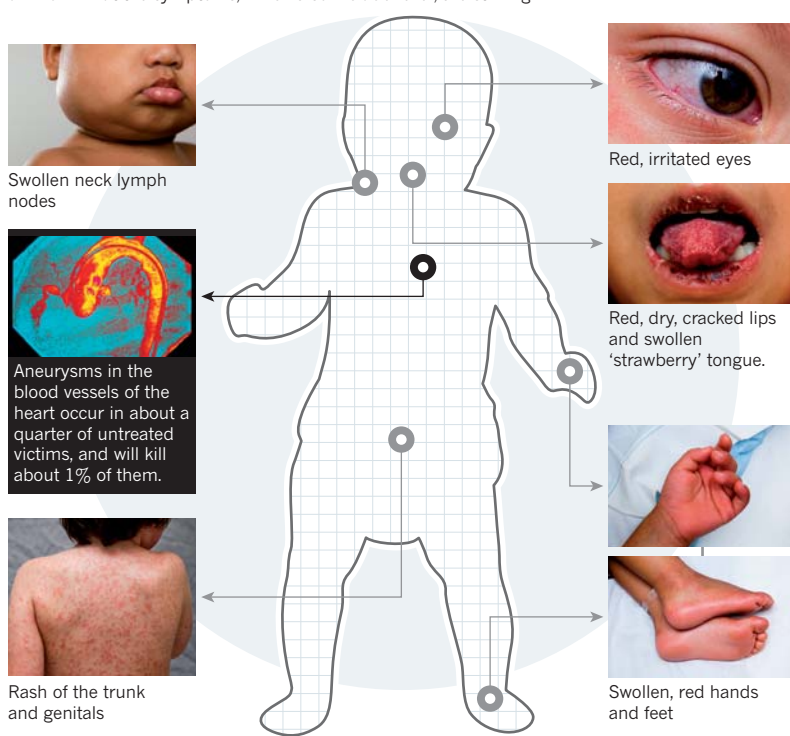
"I must say I was really shocked and surprised," says Rodó. The climate-disease correlation had fallen out with remarkable ease. And, the impli-

cation — that a human disease agent might still remain active after riding the winds all the way across the Pacific — was unprecedented.

Certainly the group's analysis would fly in the face of conventional wisdom if the disease agent proves to be a living organism, says Burns. Microbiologists have generally assumed that ultraviolet radiation and the near-cryogenic temperatures at high altitude will annihilate any infectious microbes before they can make it across an ocean. But maybe not, she says. "My background is molecular virology.

## MYSTERIOUS MALADY

Kawasaki disease, discovered in Japan in 1960, is a widespread inflammation of the body's medium-sized blood vessels that affects mainly children under the age of five. The cause is unknown. But the symptoms, which also include fever, are striking.



1980s investigating an initially promising link between Kawasaki disease and carpet cleaning, only for it to lead nowhere.

In the mid-2000s she started looking at a possible connection between Kawasaki disease and climate, along with Daniel Cayan, a meteorologist at the Scripps Institution of Oceanography at UCSD. Their first success came when they analysed the incidence of Kawasaki disease in Japan by prefecture, using Burns's contacts in the country to access detailed records of nearly 85,000 patients



When I preserve my viruses in the lab, what do I do? I desiccate them and freeze them at  $-80^{\circ}\text{C}$ . Well, hello! Those are the conditions up in the troposphere."

Besides, says Burns, wind is often full of dust. "If you take a dust particle and look at it under the electron microscope, it's like a whole universe," she says. "It's got nooks and crannies and valleys and peaks", any of which could shelter a microbe or two from solar ultraviolet.

It's certainly the case that wind can carry pathogens short distances, says Arturo Casadevall, a microbiologist at the Albert Einstein College of Medicine in New York. He points to coccidioidomycosis, or valley fever, a human fungal disease that often appears in the US Southwest after dust storms or when earthquakes shake spores from the soil into the air. And over longer distances, there is evidence that the fungus *Aspergillus sydowii* rides west on dust storms from Africa to cause disease in sea fans in the Caribbean<sup>4</sup>.

Dale Griffin, an environmental and public-health microbiologist at the US Geological Survey in Tallahassee, Florida, has studied African dust storms close up. On cruises to the waters over the Mid-Atlantic Ridge, Griffin has taken air samples just above the boat and cultured hundreds of microorganisms, including two fungal plant pathogens, the bacterium *Pseudomonas aeruginosa*, which can cause fatal infections in burn victims, and *Brevibacterium casei*, which can cause blood infections.

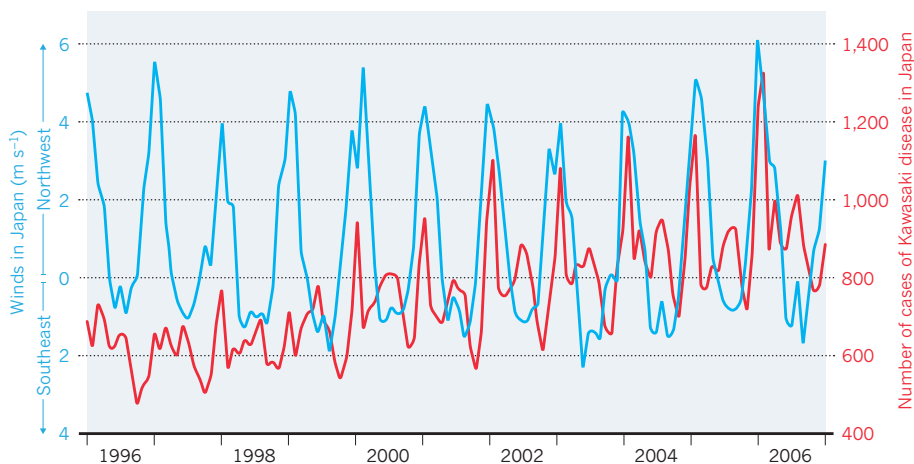
He estimates that about 10–20% of the dust-storm culture samples are pathogens. Griffin, who has also collaborated on a project that cultured microorganisms from samples collected at an altitude of 18 kilometres in the stratosphere<sup>5</sup>, believes that Rodó and his colleagues have made a strong case that Kawasaki disease is riding the wind. "It's not surprising to me that an agent that has yet to be identified could be moving in the atmosphere — with dust or without dust," he says. "We also know there are many, many species that are very tolerant and can make long journeys in the atmosphere."

#### FINDING THE CULPRIT

Casadevall agrees that the case for wind-borne Kawasaki disease is strong, but goes on to emphasize that correlation is not causation, a point also raised by researchers sceptical of the theory. "In the past, people have made claims about this sort of thing, but it's all circumstantial," says Donald Aylor, a plant pathologist at the Connecticut Agricultural Experiment Station in New Haven, who has studied the wind dispersal of plant pathogens and pollen for 35 years. Even if the wind is blowing in a certain direction when a particular disease shows up, he says, it can be difficult to prove that a wind-blown pathogen is to blame. And when the pathogen is a total unknown, as in Kawasaki disease, the difficulties are even greater, he says. "You would have to ask yourself what else was happening during the time

#### SEASONAL CYCLE

The number of Kawasaki disease cases in Japan (red) is slowly rising, for unknown reasons, but is strongly correlated with the average velocity of winds coming from the northwest (blue) — the direction of central Asia.



that the Kawasaki outbreak was happening. I mean, there could be a million things, right?"

The best way to counter this objection, Burns and her collaborators know, is to find the airborne pathogen. With Lipkin, they set up an experiment to filter the air over Japan at various altitudes during a period when the agent was suspected to be present, and then to sequence the DNA of everything on the filter — an approach known as 'metagenomics'.

In early March 2011, a Spanish engineer wearing a protective suit to prevent contamination went up in an aircraft that carried an air-filter built by Rodó's lab in Barcelona. The craft followed a route mapped out by Rodó using real-time wind data. When it returned, the samples were packed in dry ice and shipped to Lipkin's lab at Columbia. The timing was

**"IT SMELLS LIKE AN INFECTIOUS DISEASE; WE'VE JUST NEVER BEEN ABLE TO CATCH THE CULPRIT."**

particularly fortunate as the route had crisscrossed Japan's Fukushima region. Just a week later, after nuclear reactors there were damaged in the earthquake and tsunami on 11 March, the winds would be full of radioactivity.

At Columbia, where the metagenomic analysis is being carried out by biologist Brent Williams, progress has been slow because of the minuscule amounts of DNA present in air samples taken at high altitude. But the work is beginning to pay off, says Lipkin. Williams has found candidates for the Kawasaki agent — although Lipkin declined to discuss them before publication — and will soon be

progressing to immunoassays. In these, antibodies generated to proteins predicted to be expressed by the suspected disease agent will be mixed with serum samples from children who have had Kawasaki disease and from controls who have not. If the antibodies interact significantly more strongly with the Kawasaki disease samples than with the controls, the team can be more confident that its suspected Kawasaki agent is the real thing.

The next step will be to look for DNA sequences in the blood samples of affected children that match the DNA detected in the air samples. "That would also be strong circumstantial evidence that would give us confidence that we're on the right track," Lipkin says.

With the identification of a causative agent, many more questions about Kawasaki disease could be addressed. Where is the agent's natural reservoir and is there an animal host? Why did this agent only emerge to cause disease in the 1950s and 60s? And why is its incidence rising?

And, of course, what other diseases are blowing in the wind? Taiwanese scientists, noting that outbreaks of avian influenza often occur downwind of dust storms, found influenza virus in air samples, and found that concentrations of the virus spiked when dust storms blew in from central Asia<sup>6</sup>. So, says Burns, "why don't we acknowledge the possibility that agents important for human health could be travelling on these wind currents?" ■

**Jennifer Frazer** is a Colorado-based science writer and blogger for Scientific American.

1. Rodó, X. *et al. Sci. Repts* **1**, <http://dx.doi.org/10.1038/srep00152> (2011).
2. Kawasaki, T. *Arerugi* **16**, 178–222 (1967) (in Japanese).
3. Burns, J. C. *et al. Epidemiology* **16**, 220–225 (2005).
4. Shinn, E. A. *et al. Geophys. Res. Lett.* **27**, 3029–3032 (2000).
5. Smith, D. J., Griffin, D. W. & Schuergel, A. C. *Aerobiologia* **26**, 35–46 (2009).
6. Chen, P.-S. *Environ. Health Perspect.* **118**, 1211–1216 (2010).

# fMRI 2.0

BY  
KERRI  
SMITH



*Functional magnetic resonance imaging is growing from showy adolescence into a workhorse of brain imaging.*

**T**he blobs appeared 20 years ago. Two teams, one led by Seiji Ogawa at Bell Laboratories in Murray Hill, New Jersey, the other by Kenneth Kwong at Massachusetts General Hospital in Charlestown, slid a handful of volunteers into giant magnets. With their heads held still, the volunteers watched flashing lights or tensed their hands, while the research teams built the data flowing from the machines into grainy images showing parts of the brain illuminated as multicoloured blobs.

The results showed that a technique called functional magnetic resonance imaging (fMRI) could use blood as a proxy for measuring the activity of neurons — without the injection of a signal-boosting compound<sup>1,2</sup>. It was the first demonstration of fMRI as it is commonly used today, and came just months after the technique debuted — using

PADDY MILLS

a contrast agent — in humans<sup>3</sup>. Sensitive to the distinctive magnetic properties of blood that is rich in oxygen, the method shows oxygenated blood flowing to active brain regions. Unlike scanning techniques such as electroencephalography (EEG), which detects electrical activity at the skull's surface, fMRI produces measurements from deep inside the brain. It is also non-invasive, which makes it safer and more comfortable than positron emission tomography (PET), in which radioactive compounds are injected and traced as they flow around the body.

fMRI has been applied to almost every aspect of brain science since. It has shown that the brain is highly compartmentalized, with specific regions responsible for tasks such as perceiving faces<sup>4</sup> and weighing up moral responsibility<sup>5</sup>; that the resting brain is in fact humming with activity<sup>6</sup>; and that it may be possible to communicate with patients in a vegetative state by monitoring their brain activity<sup>7</sup>. In 2010, neuroscientists used fMRI in more than 1,500 published articles (see 'The rise of fMRI').

But researchers readily admit that the technique has flaws. It doesn't measure neuronal activity directly and it is blind to details such as how many neurons are firing, or whether firing in one region amplifies or dampens activity in neighbouring areas. The signal — a boost in blood flow in response to a stimulus — can be difficult to extract from the 'noise' of routine changes in blood flow, and the statistical techniques involved are easy to misunderstand and misuse. "I'm surprised that fMRI has kept going for 20 years," says Karl Friston, scientific director of University College London's neuroimaging centre. Friston says he thought all the interesting questions would have been "cherry-picked within the first two or three years".

But fMRI has kept going, in part because no other technique has bettered its ability to see what the human brain is doing. It has turned psychology "into a biological science", says Richard Frackowiak, who works with Friston. Now, scientists are intent on finding ways around some of the limitations and pushing the technique into the next 20 years. *Nature* takes a look at four futures for fMRI.

## DIRECT MEASURES

1

Perhaps the biggest conundrum in fMRI is what, exactly, the technique is measuring. Researchers know that it measures the oxygen carried

in blood by haemoglobin, and they assume that a stronger signal reflects a greater demand for oxygenated blood when neurons become electrically active in response to a task. But several papers have called this assumption into question, suggesting that blood oxygen levels could rise in preparation for neuronal activity as well as during it<sup>8</sup>; or, worse, that they could be undulating for reasons other than neuronal activity<sup>9</sup>.

Most researchers in the fMRI community are comfortable enough with the proxy to carry on doing experiments, even if not all the details have been ironed out. "We have a pretty good handle that it's measuring something that neurons are doing that's relevant to mental function," says Russell Poldrack, director of the Imaging Research Center at the University of Texas at Austin. But some teams want to do better, by getting a more direct measure of neuronal activity. "The thing that we're most interested in is not where blood flow is but where the brain is electrically active," says John George, an MRI physicist at the Los Alamos National Laboratory in New Mexico. The only ways in which electrical activity can be measured directly, however, are by placing electrodes into the brain, or by picking up electrical signals from outside the skull, a method that lacks the depth and spatial resolution of fMRI.

One solution might be to use a type of MRI that can measure the magnetic field of each neuron as it conducts electrical signals. But these perturbations are an order of magnitude smaller than those produced by changes in blood oxygen level. George's team is therefore developing a technique that uses ultrasensitive magnetometers called SQUIDs (superconducting quantum interference devices) to pick up such perturbations<sup>10</sup>. "We detect currents close to the levels we anticipate neurons would produce," he says. But the obstacles are huge. "It's very much like the early days of fMRI," says George. The next steps are to make the detection methods faster — neural signals are much quicker than those from blood — and to win over sceptics with a clear demonstration of the measurements in a tissue sample or an animal. "There are hints that signals are there, but most people don't believe it," says George. "Once they believe you can do it, they'll show you how to do it better."

## MORE THAN A PRETTY PICTURE

2

The multicoloured splodges that correspond to active brain areas have helped fMRI to earn the disparaging nickname 'blobology', reflecting some neuroscientists' frustration with the limited information that a blob conveys. It can show that a language task, for example, correlates with activity in the left hemisphere's frontal lobe, but not whether the activity is actually the result of language processing — or simply of paying attention to a screen. "You can't just infer causality from looking at where a task is happening," says Peter Bandettini, who heads the functional imaging methods section at the US National Institute of Mental Health's Laboratory of Brain and Cognition in Bethesda, Maryland. That is why the use of fMRI to show that a region is correlated with a task "is starting to slow down", he says. "No one's getting tenure based on that any more."

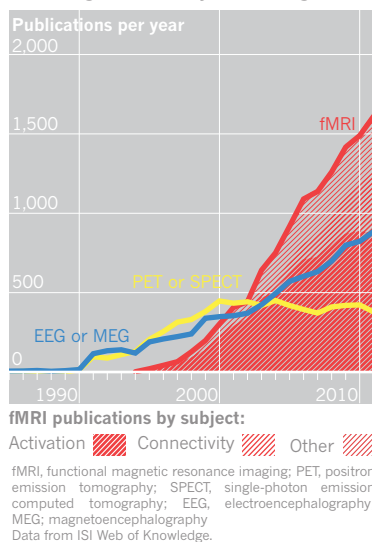
Neuroscientists are now seeking ways to build a more detailed model of the brain's organization, networks and function, so that they can interpret the patterns of activation with more confidence. A good model of brain networks might provide more detail about what happens when a person looks at a familiar face, for example, including which regions are involved in visual processing, memories and emotion; the order in which the regions respond; and how important each area is to the overall task. "The major shift is towards networks," says Stephen Smith, associate director of the Oxford University Centre for Functional MRI of the Brain, UK, whose team is working on such models. "What we're trying to get is the true underlying connectivity," he says, "rather than make a superficial comment about everything being connected to everything because they're all correlated."

A sophisticated picture of brain networks is also the goal of the Human Connectome Project<sup>11</sup>, a 5-year, US\$40-million effort funded by the US National Institutes of Health (NIH) in Bethesda, Maryland, that got under way in 2010. The project aims to map the human brain's wiring using a variety of techniques, including fMRI. Such a 'reference' connectome could help in interpreting individuals' fMRI scans and could reveal how variations in connectomes affect behaviour or contribute to disease.

Other researchers are using sophisticated statistical techniques to pick out detailed patterns from fMRI scans. One, called multivariate analysis, charts the behaviour of many units — or voxels — of brain activity in parallel, rather than averaging them together into a blob.

### THE RISE OF fMRI

Use of fMRI has rocketed, and now more studies are looking at connectivity between regions.





Blobs can identify large, active brain areas, but might miss clumps of inactive neurons within it or small islands of active neurons in quiet areas. “The more you look, the more you get meaningful information,” says Bandettini. “What previously was noise is now suddenly signal.” These techniques are even allowing researchers to work out what stimuli are present just by looking at brain activity patterns. Last year, Jack Gallant from the University of California, Berkeley, recorded the fMRI activity of three members of his lab as they watched hours of film clips. The team then developed a computational model that used fMRI scans to reconstruct a movie approximating what the people had been watching — a person wearing blue, for example, or a red bird<sup>12</sup>.

### DAMPENING THE NOISE

3

fMRI tends to generate small signals and a lot of noise. “You need quite a lot of neurons firing in synchrony with each other to see a change in blood oxygenation,” says Smith. The noise means that many changes — a small group of neurons firing together, or subtle or quick variations in oxygenated blood flow — might not be picked up. The low signal-to-noise ratio forces fMRI researchers to use statistical approaches to pick out what is significant in their scans — and that means that there are numerous ways to interpret a data set. “If you try them all, you’re going to find something,” says Poldrack.

Some groups are managing to boost the signal by using stronger magnets. In an MRI machine, a high magnetic field aligns the spins of the protons in hydrogen atoms; then radio waves knock the spins out of alignment. As the spins gradually realign, they send out a signal — or resonate — and those in areas of oxygenated blood resonate at a different frequency from those in deoxygenated blood. But only a tiny proportion of the protons react to the field and radio waves. Stronger magnets line up a greater proportion of the proton spins, which then generate a stronger signal as they realign.

The scanners used in neuroscience today typically have magnet strengths of 3 tesla, which is many thousand times stronger than Earth’s magnetic field, and have a resolution of 3 cubic millimetres. But stronger magnets are creeping into practice. In 2010, for example, scientists at the University of Nottingham, UK, used a 7-tesla magnet to build a map of the human somatosensory cortex<sup>13</sup> — which is responsible for processing touch and some aspects of movement — at a resolution of 1 cubic millimetre. The NeuroSpin facility near Paris is building an 11.7-tesla whole-body system, the strongest yet for human studies. Magnets much stronger than this cannot be used on humans, because they increase artefacts in the images and can trigger dizziness and other side effects.

Another way to increase the signal is to inject molecules that are easier to detect than oxygenated blood, in a method more akin to PET. Gary Green, director of the York Neuroimaging Centre at the University of York, UK, is working with parahydrogen, a ‘hyperpolarized’ molecule in which the proton spins are more aligned than in many other molecules, and which generates a strong signal during MRI. In 2009, Green and his colleagues showed that they could transfer spins from parahydrogen to an organic molecule without changing the latter’s chemical structure<sup>14</sup> — the first step towards preparing hyperpolarized drugs or other molecules that bind to receptors, and then track how these substances are taken up, or how they interact.

Finding better statistical ways to remove noise will also be a big help. Poldrack runs a ‘best practice’ wiki ([www.fmrimethods.org](http://www.fmrimethods.org)) that covers how

fMRI data should be analysed, and has published guidelines for how the work should be reported, recommending, for example, that researchers include all the experimental detail necessary to reproduce an analysis, such as “what your subjects were asked to do and what they actually did”<sup>15</sup>. “We need to enforce more rigour,” he says.

4

### WHICH WAY TO THE CLINIC?

Getting fMRI to the clinic is, for some, the most pressing challenge the field will face in the next few years. “It hasn’t really been used clinically yet, on individual subjects,” says Bandettini. Clinicians want to be able to ask, for example, whether a drug is working to relieve schizophrenia, or whether a person with depression is in danger of committing suicide. The difficulty lies in making sense of an individual’s scan. Most fMRI data are averages of results from many people doing the same task. This method has a higher chance of seeing a true difference between two groups or two tasks than those from an individual.

Researchers are now developing statistical methods to pull meaningful information out of a single scan. In one study<sup>16</sup> published in 2010, a team trained a computer to pick out patterns in brain-scan data collected when participants were resting. They did this for nearly 240 people aged 7–30 years to build up maps of brain connectivity at different ages. They then showed that they could take a single brain scan from a different person and, by comparing it with their reference set, work out the owner’s brain maturity. Such techniques might eventually be used to diagnose a developmental delay or psychiatric disorder, and there are hints that they can identify teenagers genetically at risk for depression<sup>17</sup>.

Having a good reference set will form the backbone of clinical fMRI, says Arthur Toga, a neurologist at the University of California, Los Angeles. Toga is a principal investigator on an effort to build such a reference, called the Alzheimer’s Disease Neuroimaging Initiative, a longitudinal study of around 800 people looking at the onset and progression of Alzheimer’s disease through genetic analyses, brain structure and function and blood biomarkers. Toga hopes that the information will form a database against which future individual scans can be compared.

With new ways both to examine the data and to boost the technology, many neuroscientists see a future filled with multicoloured blobs — albeit sharper and better-understood ones. “People will be very busy easily for the next 20 years,” says Bandettini. “I would say that fMRI in many aspects hasn’t really even begun.” ■

**Kerri Smith** is editor of the *Nature* Podcast, and is based in London.

➔ **NATURE.COM**  
Listen to a podcast  
about 20 years of  
fMRI:  
[go.nature.com/ufbguk](http://go.nature.com/ufbguk)

- Ogawa, S. *et al. Proc. Natl Acad. Sci. USA* **89**, 5951–5955 (1992).
- Kwong, K. K. *et al. Proc. Natl Acad. Sci. USA* **89**, 5675–5679 (1992).
- Belliveau, J. W. *et al. Science* **254**, 716–719 (1991).
- Kanwisher, N., McDermott, J. & Chun, M. M. *J. Neurosci.* **17**, 4302–4311 (1997).
- Greene, J. D., Somerville, R. B., Nystrom, L. E., Darley, J. M. & Cohen, J. D. *Science* **293**, 2105–2108 (2001).
- Smith, S. M. *et al. Proc. Natl Acad. Sci. USA* **106**, 13040–13045 (2009).
- Monti, M. M. *et al. N. Engl. J. Med.* **362**, 579–589 (2010).
- Sirotnin, Y. B. & Das, A. *Nature* **457**, 475–479 (2009).
- Maier, A. *et al. Nature Neurosci.* **11**, 1193–1200 (2008).
- Matlachov, A. N., Volegov, P. L., Espy, M. A., George, J. S. & Kraus, R. H. Jr *J. Magn. Reson.* **170**, 1–7 (2004).
- Bardin, J. *Nature* **483**, 394–396 (2012).
- Nishimoto, S. *et al. Curr. Biol.* **21**, 1641–1646 (2011).
- Sanchez-Panchuelo, R. M., Francis, S., Bowtell, R. & Schluppeck, D. *J. Neurophysiol.* **103**, 2544–2556 (2010).
- Adams, R. W. *et al. Science* **323**, 1708–1711 (2009).
- Poldrack, R. A. *et al. NeuroImage* **40**, 409–414 (2008).
- Dosenbach, N. U. F. *et al. Science* **329**, 1358–1361 (2010).
- Mourão-Miranda, J. *et al. PLoS ONE* **7**, e29482 (2012).

# COMMENT

**EFFICIENCY** Misplaced financial incentives are damaging science **p.29**

**FILM** The Russian microbiologist who turned revolutionary **p.34**

**MUSEUMS** The changing face of natural-history collections **p.36**

**OBITUARY** Pioneer of crystallography David Sayre **p.38**



## Clean up the waste

Fixing inefficiencies at academic institutions will strengthen — not jeopardize — teaching and research, says **Thomas Marty**.

### SUMMARY

- Academic institutions that learn to manage themselves better will achieve more with less funding in coming years.
- The main sources of inefficiencies are a wrong understanding of autonomy, weak leadership and a lack of strategic thinking when selecting research areas.
- Adapting concepts from private business will help academic institutions to address inefficiencies and get faculty members back to teaching and research.

Academic institutions are under huge pressure to do more with less — to be efficient<sup>1</sup>. Higher-education and research budgets are tight. In the United Kingdom, for instance, university funding was down by more than 12% last year. The economic downturn means that fewer families can afford tuition fees, and universities are seeing reduced financial returns on their endowments<sup>2</sup>.

Cutting costs is one way to ease this burden, but universities often gain most by producing more output with the same funding. Efficiency is largely about saving time and effort, not reducing expenditures. This frequently entails helping academic institutions to learn to manage themselves better, by adapting concepts from private business. There are differences, however: whereas unstructured time is anathema in business, it is key in research, enabling faculty members to develop new ideas.

Better management is not about telling professors how to teach and researchers how to run experiments. Cutting back on administrative and managerial inefficiencies should not conflict with the core activities of research and teaching — 'academic freedom'. Actually, the reverse is true. With my colleagues at the consultancy firm Berinform, which advises on the management of research and higher-education institutions, we have found that reducing bureaucracy can increase scholars' time for research and teaching.

### REDUCE AUTONOMY

Higher education relies heavily on the autonomous, expert work of brilliant minds. But sometimes, that autonomy can be taken too far. Some academics have a tendency to ►

ILLUSTRATION BY VIKTOR KOEN

► set their own priorities; administrative matters are regarded as unimportant and managerial decisions are usually taken at the last minute with little consideration for the consequences. Although such behaviour might be favourable to the individual, it represents a heavy burden for the institution as a whole.

We at Berinfor see many examples in which academics who insist on the wrong kind of autonomy cause a great deal of administrative waste. In one large department we worked with, each faculty member planned student courses individually. It was up to administrators to resolve time conflicts and to ensure the courses were consistent with the programme's direction. Moreover, so that students could register online, courses had to be entered in a central database, which would not recognize courses that didn't fit a certain format. Administrators were thus spending vast amounts of time mediating between the wishes of the faculty members and the needs of the central information system, a task requiring two full-time administrators.

By bringing together administrators and faculty members to get their perspectives on course planning, we helped them to agree on a standard planning process with strict deadlines for each stage, which drastically reduced the number of revisions. This freed up administrators to provide support to faculty members who were setting up a new graduate school. These efficiency gains were only possible by reducing the autonomy of faculty members, albeit in the right places. Although course descriptions had to fit the template and be handed over to administrators within the defined period, the contents were left fully to the academics.

### STRENGTHEN GOVERNANCE

A second source of inefficiency at higher-education institutions is governance. Decisions are often taken at the wrong level of hierarchy, involving too many people or too great a focus on details. This is evident in the overuse of large-membership committees, which leads to decisional paralysis and takes professors away from research and teaching.

Many committees are set up as permanent bodies with a vague remit (such as 'research committee') and are focused on consensus rather than on practical solutions to specific problems. Every hour spent on a useless committee is an hour lost for research and teaching. For example, 16 professors meeting once every 2 months for 4 hours represents about 4 days of total productive time lost each month — not counting preparation time and staff support. Taking all costs into consideration, running such a committee could amount to up to 100 working days a year. We advise our clients to review the purpose of every committee periodically, and to

assign tasks to the correct body made up of appropriate people.

Another major source of inefficiency at academic institutions is weak leadership. Faculty members often distrust leaders, particularly in European institutions. Consequently, many academic management posts, such as dean or institute director, have a short term of 2–4 years and limited executive power, ensuring that decision-making remains largely participatory. But these limitations make it difficult for leaders to bring about improvements.

Although strong-willed leaders who shape organizations to their own vision (such as Apple co-founder Steve Jobs or Jack Welch, former chairman of General Electric) enjoyed high respect in the private sector, they would encounter disdain and strong resistance in academia. A few years ago, the president of a world-class university was ousted through the political pressure of his academic peers because he wanted to improve manage-

**“Measures that are intended to increase efficiency often encounter resistance within academic institutions.”**

rial efficiency by increasing the power of department heads at the cost of the autonomy of the individual professors. When leadership is weak, reactive rather than proactive decisions prevail, and the direction of the institution barely changes. The best way to strengthen governance is to have longer terms for institute directors and university presidents, and to give them the power to make operational decisions that do not require a 20-person committee, such as how research and teaching rooms should be allocated. When an institute of one of our clients saw an exponential growth in student numbers, it increased the powers and doubled the term of the director, who made fast, executive decisions that helped to prevent problems such as oversubscribed courses.

Beyond leadership issues lies the problem of inadequate training of managers in higher education. Academics are trained to focus on one topic, whereas managers need to be generalists who can handle several diverse problems at the same time. But management is a skill that can be learned. We advise schools to invest in the management skills of their academic directors, which may require as little as an intensive week-long course.

### THINK STRATEGICALLY

When we advise institutions on how to improve research efficiencies, we don't try to change what they decide to study. But we do advise them to keep some things in mind when they make that decision. Notably, doing more in a given research area increases the efficiency of the whole institution.

The more people that work on one topic together, the more they discuss their ideas and techniques, focus their efforts and avoid duplication. This holds true for expensive infrastructure: doing more experiments on a synchrotron or electron microscope, for example, improves their output because later experiments profit from the learning invested in earlier ones. But too often, institutions aim for breadth, hiring faculty members who study a wide range of topics, which prevents these economies of scale.

Departments also tend to hire academics who are studying hot topics, such as stem cells. But not everyone can bring together the critical mass of scientists required to become a prominent and successful stem-cell institute. Instead, we advise clients to consider their positioning in scientific fields — to identify and expand areas in which they are already doing well, rather than starting small institutes from scratch. The bigger the centre for a particular subject, the more likely it is to attract brilliant minds. For example, an institution with a strong backbone in physical chemistry, experimental physics and engineering might bring those fields together to build a world-class materials-science department.

Only by addressing all three of the areas I have identified — autonomy, governance and strategy — will institutions fully address internal inefficiencies. Despite obvious benefits, measures that are intended to increase efficiency often encounter resistance within academic institutions, especially if they are proposed by outside actors such as governments or advisory groups<sup>3</sup>. However, none of these suggestions would threaten academic freedom. Although many of these recommendations are based on concepts that come from the private business sector, we at Berinfor do not support replacing the academic culture with a business mindset<sup>4</sup>, which would reduce both creativity and productivity. Instead, higher-education institutions should develop their own culture — including needs shaped by academics, administrators and leaders — and translate business concepts to fit the academic environment. ■

**Thomas Marty** is a senior consultant at Berinfor, a consulting firm in Zurich, Switzerland, specializing in the management of academic institutions.

e-mail: [thomas.marty@berinfor.ch](mailto:thomas.marty@berinfor.ch)

1. *Efficiency and Effectiveness in Higher Education* (Universities UK, 2011); available at <http://go.nature.com/bnmmnb>
2. Lewin, T. *New York Times* (14 November 2009); available at <http://go.nature.com/1x4byp>
3. Diehl, R. L. & the Executive Leadership Team *Maintaining Excellence and Efficiency at the University of Texas at Austin* (College of Liberal Arts, 2011); available at <http://go.nature.com/evhmqy>
4. Balocco, M., Licka, P. & Marty, T. *Wissenschaftsmanagement* 4/11 (2011) (in German).





# Perverse incentives

Counterproductive financial incentives divert time and resources from the scientific enterprise. We should spend the money more wisely, says **Paula Stephan**.

Scientists may portray themselves as not being motivated by money, but they and the institutions where they work respond in spades to financial opportunities. Incentives that encourage people to make one decision instead of another for monetary reasons play an important part in science. This is good news if the incentives are right. But if they are not, they can cause considerable damage to the scientific enterprise.

For instance, cash incentives adopted by countries such as China, South Korea and Turkey encourage local scientists to submit papers to high-end journals despite the low probability of success. These payments have achieved little more than overloading reviewers, taking them away from their work, and have increased submissions by the three countries to the journal *Science* by 46% in recent years, with no corresponding

increase in the number of publications<sup>1</sup>.

Sadly, science is full of incentives gone awry. Look no further than expanding PhD programmes that produce graduates with almost no career prospects, or the growth of lab space with no apparent increase in productivity.

The economic rules behind science were written without much consideration for unintended consequences, but such consequences abound because people and institutions are so responsive to incentives. And in the current economic climate, no one can afford to waste time or resources. In a world of tight budgets, getting the incentives right is more important than ever.

## BAD DIRECTIONS

Consider the financial calculations that encourage universities to hire a series of postdocs rather than staff scientists. Postdocs earn around half to two-thirds of a staff scientist's salary. They are young, have fresh perspectives and new ideas and are temporary, so can be let go when budgets decline<sup>2</sup>. But, in reality, postdocs are not cheap: substantial resources — both their own and society's — have been invested in training them.

If a postdoc doesn't get a research job, taxpayers do not get a return on their investment. Neither does the postdoc: someone who did not go to graduate school and entered the labour market in 2001 was earning about US\$58,000 in 2008; a first-year postdoc who started graduate school in the United States in 2001 was making around \$37,000 in 2008 on graduation<sup>3</sup>. During a three-year postdoc position, a scientist gives up more than \$60,000 on average in return for highly uncertain job prospects. And many postdocs will not get a research job. There are few faculty openings, and limited numbers of research positions in government and industry. So even if individual postdocs cost less, from a societal perspective they can be expensive.

Equally harmful are rules that encourage scientists to support graduate students on a research assistantship (RA) rather than on a training grant, despite evidence that the ►

## SUMMARY

- Science is full of incentives that encourage bad financial choices, such as expanding labs and hiring too many temporary scientists.
- These incentives hurt both individual scientists and society as a whole, which gets minimal return on its investment when someone is trained for a field with no career prospects.
- The way forward is to fix incentives that are damaging the system, by considering their true social and personal cost.

► latter produces better outcomes. Part of the reason is that RA funding comes with an additional amount to cover the university's overhead, or indirect rate, which may be as high as 50%. For those on training grants from the US National Institutes of Health (NIH) in Bethesda, Maryland, that amount is capped at 8%. This difference easily translates into an institution getting at least \$12,000 more for every RA-supported student. Other considerations affect the choice of RAs over training grants, too — RAs are under the direct control of principal investigators, whereas graduate students on training grants are less so.

However, training grants are arguably better for scientists in the long term. Importantly, they give departments the incentive to provide a high-quality training experience, because renewals for training-grant awards are evaluated on the quality of the PhD experience and placement outcomes. By contrast, scientists who support students on research grants are not required on renewal to disclose where graduates end up being placed. Some principal investigators collect this information, but departments typically do not — my informal survey of 45 science departments found only two that were able to report where their students had been placed. Without this knowledge, prospective students will not be able to judge whether a lab is a good place to begin a successful science career.

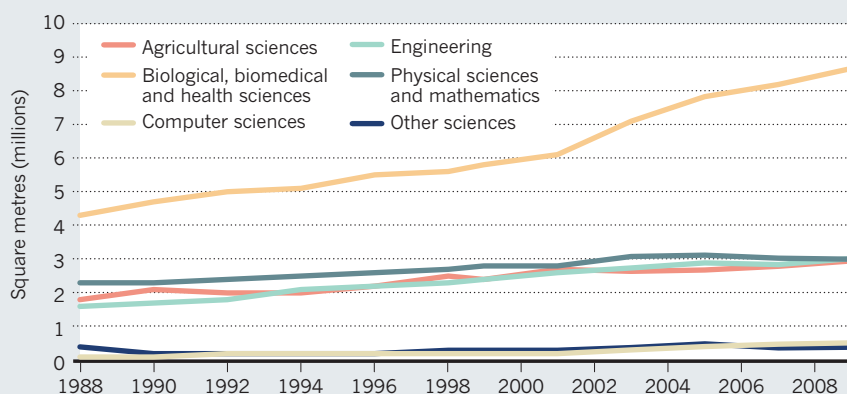
The growth of labs is another result of incentives. Bigger is seen as better: more funding, more papers, more citations and more trainees — regardless of whether the market can sustain their employment. Some institutions pay bonuses to faculty members on the basis of the amount of external funding they receive<sup>4</sup>. But, again, too many trainees creates a glut of people who will not find suitable jobs. It would have been more efficient for both the students and society to steer them in a different direction. And big labs can be wasteful — an analysis by the US National Institute of General Medical Sciences in Bethesda, Maryland, found that an increase in funding is not associated with a substantial increase in output when measured by the number of grant-linked publications<sup>5</sup>.

Other economic incentives indirectly render the scientific process less efficient — such as the tendency of scientists to avoid risk by submitting to funding organizations only those proposals that they consider 'sure bets'. This tendency comes from the need for faculty members to obtain grants to support their salary, the emphasis on preliminary data in grant applications and the difficulty of obtaining funding in today's climate.

***"The building boom is now costing the scientific enterprise by creating space that cannot be paid for."***

## BIOLOGY'S RESEARCH FOOTPRINT

There has been huge expansion in laboratory space for the biomedical sciences since the late 1990s in comparison with other fields.



SOURCE: US NATL SCI. FOUND.

If most scientists are risk-averse, there is little chance that transformative research will occur, leading to significant returns from investments in research and development. Funding bodies sometimes give money specifically for field-changing research, but not nearly enough — Pioneer grants from the NIH fund fewer than 1% of applicants.

In the European Union, there are strong incentives for researchers to team up with colleagues in other countries. This is because most funding opportunities under the various research Framework programmes require consortia to be made up of at least three entities in three different European countries. No collaboration, no grant. Is this a good use of resources? Although there is evidence that collaboration leads to better research, I do not know of any that supports the idea that those collaborators should come from different countries. The extra costs of coordination — organizing the work, travel, meetings and so on — can be large relative to the money being invested in research.

Universities are also driven by incentives. By hiring faculty members on 'soft' money, with grants providing the salary, the institutions bear almost none of the risks. Furthermore, universities prefer to put up a new building or invest millions in remodelling existing lab space rather than house scientists in older buildings that they already own. Why? One reason is that debt can be an accounting asset. A US government accounting rule called A21 means that the more debt universities have from construction, the more they can add to grants for overhead costs. If a university borrows \$100 million to build a new facility and pays 4% interest, it can increase its indirect rate by including the \$4-million interest payment in the calculation. The building binge is further fuelled by competition among universities: recruiting senior faculty members requires space, and lots of it.

What is so bad about institutions putting up new research facilities? The answer lies

in what economists call 'incidence' — who eventually pays. Before the global financial crisis, universities had hoped to recoup the money through increased indirect costs and through the 'buy-out' money they receive from funding agencies to cover the salaries of permanent faculty members working on grants. But now that grants are harder to get, that money isn't coming in. Unless states and private institutions default, someone will have to pay the bonds. The money is likely to come from the physical sciences, the humanities and social sciences, as well as cutbacks in hiring across departments. In short, the building boom is now costing the scientific enterprise by creating excess space that cannot be paid for.

## FIX WHAT'S BROKEN

The way forward is to alter these damaging incentives. The scientific enterprise should cut back on the demand for graduate research assistants by establishing more research institutes that are not focused on the production of PhDs, such as the Howard Hughes Medical Institute's Janelia Farm campus in Ashburn, Virginia. Research institutes, by producing fewer PhDs, lead to a better balance between supply and the limited number of research jobs. Abstinence, after all, is the most effective form of birth control.

In addition, we should consider ways of making graduate students and postdocs more costly to universities, to discourage their overuse and reflect their social cost. One possibility is to 'tax' the two positions, making them more expensive relative to other staff types, thereby providing an incentive to employ permanent rather than temporary staff. Principal investigators and their departments should also be required to report placement data online as part of all research-grant applications. This would allow society to monitor the return on its investment, and students to assess job outcomes.

Training grants should be made more



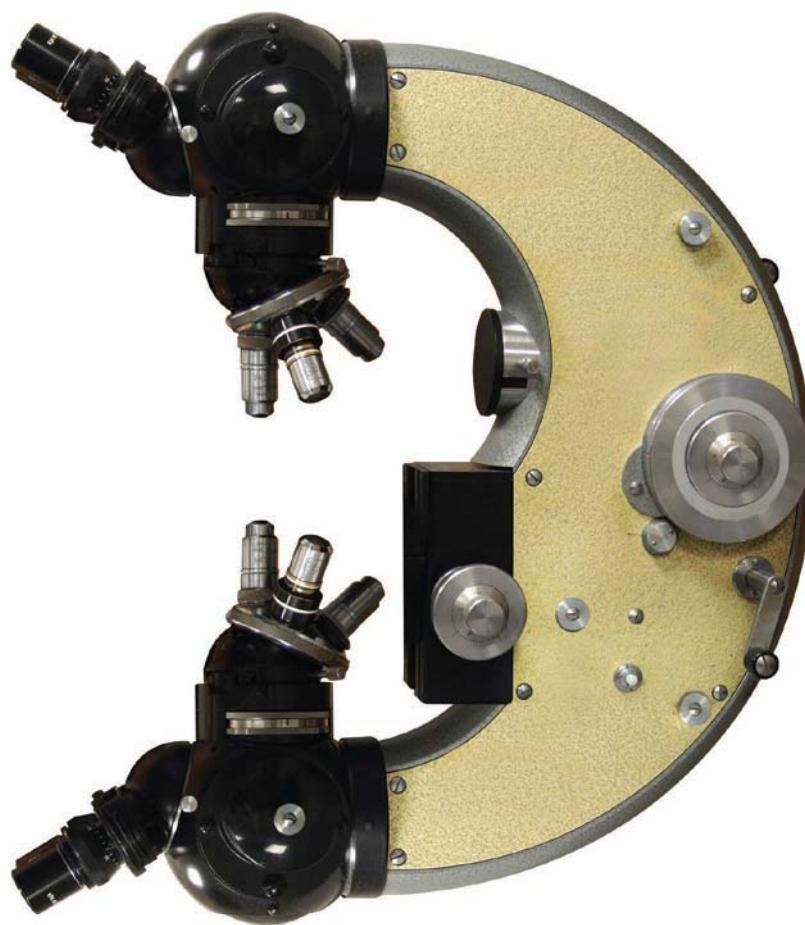
economically attractive. And rules should be altered to limit the amount of interest payments universities can include when calculating indirect rates, and the amount of faculty members' salaries that can be charged to grants, thereby dulling the incentive to hire people for soft-money positions. Shifting evaluations from projects to people, and de-emphasizing the importance of metrics in hiring and promotion, could encourage scientists to work on riskier projects<sup>6</sup>.

Many of the problems now faced by science accelerated when biomedical funding started to increase steeply. For instance, the doubling of the NIH budget from 1998 to 2003 triggered universities to hire more people and build more buildings, while scientists increased the number of grants they submitted and the size of their labs (see 'Biology's research footprint'). Now, this biomedical machine needs increasing amounts of money to sustain itself — larger labs need more grants, which leads to lower success rates, with calls for more funding.

Biomedical research has done much to contribute to increased life expectancy. But it seems likely that diminishing returns have set in. New drugs are slower in coming to market and there was a less than stellar increase in US publications associated with the NIH doubling<sup>7</sup>. Moreover, many of the breakthroughs that have contributed to better health outcomes have come from other fields of science — such as the laser and magnetic resonance imaging. Funds for the physical sciences in the United States (in terms of the percentage of federal research funding) are close to a 35-year low. Perhaps it is time for deans in the biomedical sciences to rent some of that excess space to their colleagues in chemistry and physics. ■

**Paula Stephan** is professor of economics at Georgia State University in Atlanta, Georgia 30303, USA, and a research associate at the National Bureau of Economic Research in Cambridge, Massachusetts. She is author of *How Economics Shapes Science* (Harvard Univ. Press, 2012).  
e-mail: pstephan@gsu.edu

1. Franzoni, C., Scellato, G. & Stephan, P. *Science* **333**, 702–703 (2011).
2. Stephan, P. *How Economics Shapes Science* (Harvard University Press, 2012).
3. Current Population Survey, Table P-28 (United States Census Bureau, 2010); available at <http://go.nature.com/neix5i>
4. Mallon, W. T. & Korn, D. *Science* **303**, 476–477 (2004).
5. Berg, J. 'Another Look at Measuring the Scientific Output and Impact of NIGMS Grants' (22 November 2010); available at <http://go.nature.com/dae21z>
6. Azoulay, P., Graff Zivin, J. S. & Manso, G. *Incentives and Creativity: Evidence from the Academic Life Sciences* (2010); available at <http://go.nature.com/biey9d>
7. Sacks, F. *The Scientist* (11 September 2007); available at <http://go.nature.com/bpa3xr>



## Turn the scientific method on ourselves

How can we know whether funding models for research work? By relentlessly testing them using randomized controlled trials, says **Pierre Azoulay**.

**I**n times of tight budget constraints, scientists' wranglings about the real and perceived sins of public funding agencies become particularly acute. Complaints usually lead to the creation of a panel of respected, thoughtful and well-meaning scientists who come up with a plan of reform based on their intuition and experience. Funding agencies, who are genuinely concerned about improving the productivity of the scientific enterprise, often adopt these recommendations, at least in part. In one example of this process, the US National Institutes of Health (NIH) in Bethesda, Maryland, has created a large array of funding mechanisms,

each one targeted to a particular problem — including the K99/R00 or 'kangaroo' grants, which pair postdoctoral scientists with mentors to help them to prepare for tenure-track faculty positions and funding independence. Not only is this range of mechanisms confusing and costly to administer, but the effectiveness of such reforms is never seriously evaluated.

It is time to turn the scientific method on ourselves. In our attempts to reform the institutions of science, we should adhere to the same empirical standards that we insist on when evaluating research results. We already know how: by subjecting proposed reforms to a prospective, randomized ▶

ILLUSTRATION BY VIKTOR KOEN



► controlled experiment. Retrospective analyses using selected samples are often little more than veiled attempts to justify past choices.

What could such a formal experiment look like? Let me give an example. It is well documented that the past 30 years have seen a marked increase in the age at which academic scientists achieve funding independence<sup>1</sup>. One way to ensure the continued injection of talent into these ranks would be to evaluate first-time applicants separately from a larger pool, and dedicate to them a predetermined share of the available funding. An alternative would be to keep the current system in place, but to award their proposals 'bonus points'. Which reform should we adopt? And what if the 'greying' of the scientific workforce does not stem from institutional failure, but rather reflects the influence of an ever-expanding burden of knowledge, whereby scientists must spend more time in training before they can become productive<sup>2</sup>?

To test these questions empirically, for example within the NIH (the agency I have studied most closely), we could choose a random subset of funding panels to implement the first method and a second subset to implement the other. A third subset, in which funding panels proceed with business as usual, would serve as the control group in the experiment. Ideally, the study would be designed to avoid 'panel shopping' by applicants; the 100,000 or so R01 grant proposals reviewed each year by the 183 NIH funding panels are more than enough to craft a statistical protocol with adequate power.

These experiments could exist outside government agencies, too. When philanthropic organizations develop new models to fund research, they should formally investigate how their approach compares with the dominant model, which averages experts' scores to determine the funding-priority ranking for particular projects. Some emerging models, for example, give higher ranking to projects that elicit enthusiasm and controversy than to projects that generate more consensus but only tepid support across reviewers. It might be that funding a project on the basis of reviewer sentiment is more likely to result in the selection of truly innovative, field-changing projects, but how will we know for sure? A serious evaluation of this question would compare the two systems by randomizing proposals to one of these two ranking approaches, and then examining which portfolio of projects is most successful.

When I suggest these experiments,

*"Experimenting on ourselves may well lay bare some shortcomings of the scientific community."*

I encounter a lot of resistance. Wouldn't this be gambling with scientists' careers? How can we measure success — by counting publications and citations, looking at the students trained as a by-product of these grants, or using other metrics? Won't this work shift scarce funding away from actual scientific investigations?

These criticisms are without merit. The current system already gambles with scientific careers, just in a haphazard way. Scientists often disagree on how to measure success, and the choice of metric, as well as the period necessary for a careful assessment, will always be context-dependent and controversial. With the good will of administrators in public agencies or private foundations, these experiments could be rolled out with minimal disruption for about the cost of a R01 grant (typically US\$250,000 per year for 5 years). If the scientific community could test even a small number of hypotheses in this way, the system-wide benefits would dwarf this modest investment.

I am well aware that this vision will sound utopian to some. Sceptics abounded when my colleagues at the Massachusetts Institute of Technology in Cambridge founded the Jameel Poverty Action Lab and subjected development-assistance methods to randomized, controlled trials to see which worked best. But as a result of this work, policy-makers now know that it is better to give out free mosquito nets to prevent malaria than to charge even a low price for their purchase — one example among many<sup>3,4</sup>.

We inherited the current institutions of science from the period just after the Second World War. It would be a fortuitous coincidence if the systems that served us so well in the twentieth century were equally adapted to twenty-first-century needs. Experimenting on ourselves may well lay bare some shortcomings of the scientific community and expose us to criticisms from politicians, who are always looking for excuses to cut science funding. But the only alternative to such controlled experimentation is the gradual stultification of our most cherished scientific institutions. ■

**Pierre Azoulay** is an associate professor in the Massachusetts Institute of Technology Sloan School of Management, and a research associate at the National Bureau of Economic Research, Cambridge, Massachusetts, USA. e-mail: pazoulay@mit.edu

1. Rockey, S. Age Distribution of NIH Principal Investigators and Medical School Faculty (13 February 2012); available at <http://go.nature.com/2uzqca>
2. Jones, B. F. *Rev. Econ. Stud.* **76**, 283–317 (2009).
3. Cohen, J. & Dupas, P. Q. *J. Econ.* **125**, 1–45 (2010).
4. Banerjee, A. V. & Duflo, E. *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty* (PublicAffairs, 2011).



The pyramids at Giza — the culmination of two-and-a-half millennia of developing civilization — have tantalized archaeologists for centuries.

## ARCHAEOLOGY

# From ploughs to pyramids

**Andrew Robinson** discovers gems in a grand overview of ancient Egypt and the life of a pioneer in Egyptology.

**T**he Great Pyramid at Giza was once thought to be geometrically perfect — not least by Isaac Newton. It wasn't until 1798, when Napoleon Bonaparte's surveyors arrived in Egypt, that the pyramid was discovered to be flawed. Even so, the deviation from a perfect square in the four final baselines of this extraordinary edifice, constructed as a tomb for pharaoh Khufu (also known as Cheops) around 2550 BC, is less than 18 centimetres in 230 metres.

How this precision was achieved is now well understood. But mystery shrouds Khufu himself: all that Napoleon's scientists found in his burial chamber was an empty sarcophagus. It had been robbed in antiquity. As Egyptologist John Romer writes in *A History of Ancient Egypt: From the First Farmers to the Great Pyramid*, "Nothing of Khufu the man is known."

Despite many such historical blanks, Romer aims to provide a monumental survey of two and a half millennia, from Egypt's agricultural dawn in 5,000–4,000 BC to the reign of Khufu. This is the first volume of a projected two-volume study — the fruit of the author's archaeological activity in Egypt since the mid-1960s.

Romer's book is the most wide-ranging since British philologist Alan Gardiner published *Egypt of the Pharaohs* (Oxford Univ. Press, 1961), which itself replaced US archaeologist James Henry Breasted's *A History of Egypt* (Scribner, 1905). Breasted's study

## **A History of Ancient Egypt: From the First Farmers to the Great Pyramid**

JOHN ROMER

Allen Lane: 2012. 512 pp. £25

## **American Egyptologist: The Life of James Henry Breasted and the Creation of His Oriental Institute**

JEFFREY ABT

University of Chicago Press: 2012. 536 pp. £29, \$45

was a popular hit that was also admired by Egyptologists, and it is fitting that he should at long last get his due in Jeffrey Abt's deeply researched biography, *American Egyptologist*.

Romer is fortunate in that a mass of important archaeological discoveries have been made in the past 50 years. In the 1980s, for instance, German archaeologists clearing a royal tomb at Abydos discovered the oldest group of inscribed artefacts so far known in Egypt, dating to about 3200 BC. They include nearly 200 bone and ivory tags, once probably attached to grave goods. At least some of the pictograms on the tags strongly resemble the first hieroglyphs, which date from the early third millennium BC.

Earlier finds also feature: Romer describes an underground tomb in the shadow of the Great Pyramid, discovered in 1925. Its owner,

Hetepheres, was identified from damaged golden hieroglyphs originally mounted on strips of ebony on

the back of a small royal palanquin. One of her titles was "mother of the King of Upper and Lower Egypt"; another identified her as a wife of Sneferu, Khufu's father, making her a candidate for the pyramid-builder's mother. Hetepheres's sarcophagus was as empty as her putative son's, but four packets of her entrails — some still steeping in natron — were discovered in an alabaster chest. As Romer says, it is "precisely this mixture of intimacy, anonymity and grandeur, at once alien and familiar" that makes Egyptology so fascinating.

However, the lack of data poses a challenge to historians of ancient Egypt. Romer breaks away from the approach of Breasted and Gardiner, who understandably tended to emphasize the knowledge derivable from Egyptian scripts and to downplay non-literate sources. Unlike Romer, they also concentrated on the ruling elite, and tended to look through the lens of their training in the history of ancient Greece, Rome and the Bible lands.

Breasted (1865–1935) has several claims to fame apart from his *History*. He was the first US citizen to gain a PhD in Egyptology. He assiduously documented hieroglyphic inscriptions using photographs, and played an important part in reading the inscriptions in the tomb of Tutankhamun. He founded the influential Oriental Institute at the University of Chicago in Illinois in 1931. He coined the phrase 'fertile crescent' to refer to the area of the Middle East — chiefly Mesopotamia — in which some of the earliest

C. MILLER/ALAMY

**NATURE.COM**  
For more on modern  
Egyptology:  
[go.nature.com/eqwdoh](http://go.nature.com/eqwdoh)



► human civilizations arose. And some have suggested that he was the model for the fictional archaeologist Indiana Jones, who studied at the University of Chicago. In 1931, Breasted's portrait featured on the cover of *Time* magazine.

Breasted's weakness as a historian was his religious motivation, as Abt clearly, if tactfully, reveals. Although hardly a Christian fundamentalist, Breasted saw Egypt as "an early example of mankind's potential but, as it had not been Christian ... bound to fail", according to a single dismissive sentence in Romer's book.

Romer's book is not without weaknesses of its own. His stated intent is "first, to provide a guide through the intellectual quagmire that modern 'ancient Egypt' has become, and also, by careful observation, to set some of the things the ancient people made back into their original realities." Although successful in his second aim, he is perhaps less so in his first.

It is not easy to enliven prehistory while simultaneously respecting limited archaeological evidence and avoiding novelistic pitfalls. But Romer manages it: his first chapter is a sparkling evocation of how, under the influence of the River Nile, Egypt's earliest inhabitants turned from hunter-gathering to farming around Lake Faiyum, in a depression in the Sahara. He shows how flint arrowheads found at Faiyum became less efficient, yet finer, over time, showing the growth of aesthetic feeling that would result in intricate weaving, elegant pots, beautiful paintings, expressive hieroglyphs and gigantic pyramids.

However, as a guide through the "intellectual quagmire", the book is sometimes tendentious in its anxiety to avoid conventional views on the enigmatic origin of the hieroglyphs. On aesthetic grounds, Romer assumes that the Egyptians borrowed from the writing system of Uruk in Mesopotamia (invented around 3300 BC), but this is not as obvious as he says: hieroglyphs could have been an indigenous invention. Moreover, the connection between the hieroglyphs and the simple pictograms found in Abydos is not clear: the pictograms do not all face to the right, like typical hieroglyphs.

But it would be churlish to suggest that Romer's project to shed light on an Egyptian era where historians fear to look is less than welcome. After a long wait, we have an up-to-date, stimulating account of the birth of what may turn out to be the world's oldest civilization. ■

**Andrew Robinson** is based in London and is the author of *Cracking the Egyptian Code: The Revolutionary Life of Jean-François Champollion* (Thames and Hudson, 2012).  
e-mail: andrew.robinson33@virgin.net



Sergej Tschachotin, pictured in 1907, used his ultraviolet scalpel to study sea-urchin parthenogenesis.

FILM

# A radical in the lab

**Alison Abbott** enjoys the story of a cell biologist whose incendiary life was shaped by revolution.

A distracting weakness for women and politics meant that, unlike his friends Albert Einstein and Ivan Pavlov, the Russian microbiologist Sergej Stepanowitsch Tschachotin (1883–1973) never achieved scientific immortality. But in his time he was a renowned cancer researcher and innovative cell biologist — and his extraordinary life makes a riveting film.

Married five times and a frequent political exile, Sergej Tschachotin lived a life that reflected the tumult of twentieth-century Europe. So German film-maker Boris Hars-Tschachotin, one of Sergej's roughly 27 great-grandchildren, had a family story ripe for documentary. Researching tales told by his grandfather Wenja Tschachotin, one of Sergej's eight sons, Boris made an unexpected discovery. Hidden in the Paris home of his great-uncle Eugen Tschachotin sat the urn containing Sergej's ashes, which had been collecting dust for 30 years. The film *Sergej in the Urn*, out this year in Germany, describes Boris's efforts to bring the family together to fulfil Sergej's last wish: to be buried in Corsica.

Irresolvable hostilities among Sergej's four remaining sons halted that plan. But the documentary is all the more absorbing for that. It reconstructs the scientist's life and times

from filial memories, archival footage of Nazi Germany and revolutionary Russia, and an unpublished autobiography that Eugen smuggled out of the Soviet Union with the urn, letters and other documents.

Sergej's dual life as idealist-activist and scientist began early. As a student in Moscow, he was deported in 1902 for participating in protests against the Tsarist regime. He moved to Germany, where he continued his studies in Munich, Berlin and finally Heidelberg.

His first academic post was in Messina, Italy, where he was caught in the 1908 earthquake that killed more than 70,000 people. Miraculously, he and his first wife and son were pulled from the rubble alive, although Sergej's leg was crushed. In his autobiography he claims that during the subsequent surgery he thought up the 'ultraviolet scalpel' — a beam of ultraviolet light that can make precise cuts through biological tissue — for which he became moderately famous.

Back in Heidelberg, he made a prototype. He used it to manipulate individual cells under the microscope, which helped him to contribute to debates on subjects such as the response of cancer cells to treatments. He also used the scalpel to prod unfertilized sea-urchin eggs into undergoing parthenogenesis.

On the strength of these successes, and to his enormous pride, Sergej was invited to join

**Sergej in the Urn**  
DIRECTED BY BORIS  
HARS-TSCHACHOTIN.  
*Liquid Blues*  
Production: 2009.  
DVD release August  
2012.

LIQUID BLUES PRODUCTION

**"Sergej in the Urn shows a man whose family life was incompatible with science and revolution."**



Pavlov's lab in St Petersburg in 1913. But unable to resist the siren call of politics, he began to educate soldiers in science and technology matters relevant to the war against Germany. Pavlov asked him to do this outside the lab, and Sergej drifted away from research.

For the next 15 years, politics dominated his life, and his fiery reputation lost him several academic jobs. He briefly joined the White Army, and fled Russia after the 1917 Bolshevik revolution. In 1922, Sergej became reconciled with the Soviet government, and found work at the Soviet embassy in Berlin, where he befriended Einstein. But a few years later, when Stalin rose to power, Sergej gave up on the Soviets and went back to science.

His next post was in Genoa, Italy. It was supported by a grant from the Vatican, which apparently thought that his work on parthenogenesis might provide insight into the conception of Jesus.

In 1930, Einstein recommended Sergej for a post at the Kaiser Wilhelm Institute for Medical Research in Heidelberg. But Sergej was thrown out of the institute in 1933 for co-founding an anti-Nazi movement called the Iron Front. He believed in fighting propaganda on its own emotive terms, and designed a strong logo to rival the swastika and a raised-fist salute to match the raised hand.

Exiled in Paris, he explicated his propaganda theory, which drew on Pavlov's theory of conditioned reflex, in his still-admired 1939 book *The Rape of the Masses*. In 1941, with the Nazis occupying the city, Sergej was interned for some months — but pressure from German scientists secured his release.

After the war, Sergej campaigned against the atomic bomb, founding the Science Liberation Movement in Paris in 1946. He returned to his homeland in 1958, after Stalin's death. Sergej was by then into his seventies, and his letters describe his disappointment with the revolution. He was forbidden to travel, and ended his days working at the Moscow Institute of Developmental Biology. (However, he managed to find enough energy to marry for a fifth time.)

*Sergej in the Urn* shows a man whose family life was incompatible with science and revolution. This is poignantly clear when Eugen plaintively asks, "He always wanted to save the world, but what becomes of us?" Towards the end of the film, Eugen's treatment of his great-nephew shows the psychological toll of Sergej's behaviour, but to say more would be a spoiler.

*Sergej in the Urn* won best German-language documentary at the Munich Documentary Film Festival in 2010, and went on general release in Germany on 23 February 2012. The DVD (in German, with English subtitles) will be available through [www.sergej-in-der-urne.de](http://www.sergej-in-der-urne.de) from August. ■

**Alison Abbott** is Nature's Senior European correspondent.

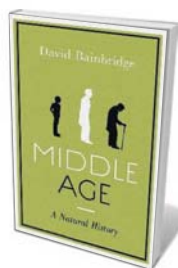
## Books in brief



### How It Began: A Time-Traveler's Guide to the Universe

Chris Impey NORTON 448 pp. \$27.95 (2012)

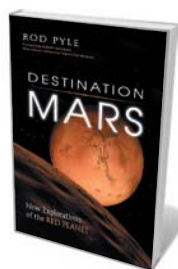
Astronomer Chris Impey takes us on a celestial road trip into deep space and time. His guided tour of the Universe starts with Earth's near neighbours — from the Moon to the star Proxima Centauri, the Orion nebula and the Milky Way — and journeys to the farthest edges of the cosmos and the first star. He pulls up at the ultimate grey area: the infant Universe, and the cosmological speculation about it. Each leg of the trip packs in science, history and anecdote, and is topped and tailed with imagined descriptions of each starry port of call.



### Middle Age: A Natural History

David Bainbridge PORTOBELLO BOOKS 304 pp. £14.99 (2012)

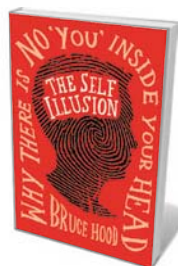
Turning 40 can be a time for celebration — or anguish. David Bainbridge, a veterinary surgeon with a penchant for evolutionary zoology, passed that watershed seething with curiosity about middle age. Sifting findings from anthropology, neuroscience, biology and psychology, he intelligently tackles tough issues such as whether there is a 'clock of death' — a genetically programmed march to oblivion. He concludes that middle age is a peak, not a slide: a distinctly human, built-in condition characterized by energy efficiency, mental stability, productivity and massive potential.



### Destination Mars: New Explorations of the Red Planet

Rod Pyle PROMETHEUS 280 pp. \$19 (2012)

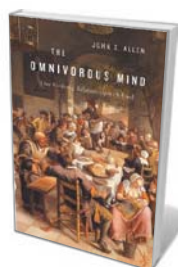
The seductive fascination of the red planet never palls, and science writer and documentary maker Rod Pyle stokes our hunger. For the Mars obsessed, the real thrills will be in his detailed descriptions of upcoming missions, the pseudo-Martian research conducted in Earth's most hostile environments, and interviews with explorers such as Steven Squyres, principal investigator of NASA's Mars Exploration Rover. Pyle's look at the planet and our perceptions and probings of it also covers Mars's geography, geology and hydrology, and its cultural history on Earth.



### The Self Illusion: Why There is No 'You' Inside Your Head

Bruce Hood CONSTABLE 272 pp. £12.99 (2012)

Day to day, we experience a sense of self but, says Bruce Hood, it is a fabrication generated by our brains. Furthermore, that sense is symphonically distributed — created by a range of brain processes rather than centred in one site. Director of the Bristol Cognitive Development Centre, UK, Hood has amassed a mountain of support for his argument — covering brain development through social interaction such as attachment, the importance of social mimicry, the illogicality of free will, online and offline 'selves' and much, much more.



### The Omnivorous Mind: Our Evolving Relationship with Food

John S. Allen HARVARD UNIVERSITY PRESS 328 pp. \$29.95 (2012)

Whether we're obsessing over intricate recipes or daydreaming about chocolate, our minds are often focused on food. Neuroanthropologist John Allen uses this mental gustation as a lens on our biological and cultural past, through anthropology, food history and the experience of chefs. The result is a banquet. Ranging over food cravings and aversions, cultural preferences and diets, he serves up plenty of *amuse-bouches*, not least an unusual take on the global love for the crispy and crunchy.



Berlin's natural history museum has exploited the aesthetic qualities of its spirit collection.

## MUSEUMS

# A natural evolution

Henry Nicholls reflects on how the 'phenotypes' of natural history museums are adapting to change.

How are the world's natural history museums faring in the twenty-first century? There can be little doubt that humanity is obsessed with nature. If you were to pile up every painting, potsherd and preserved bit of flora or fauna on show in every UK museum, for instance, around half of the artefacts would be objects of natural history. But in recent years, research practice and priorities have undergone a rapid evolution that exhibitions have struggled to reflect.

The gradual disappearance of empire-building aspirations, the emergence of the conservation movement and the rise of molecular biology have led to a decline in the collection of whole organisms. Fossils and minerals remain fair game, but researchers are mostly content to leave plants and animals where they are; a tissue or DNA sample is enough to keep them busy for decades. Yet it is still common for visitors to natural history museums to be confronted with lots of taxidermy and little interpretation.

Bringing front-of-house exhibitions into line with behind-the-scenes research is crucial. With climate change and loss of biodiversity a reality, and conflict over food production and water likely, natural history research is of increasing relevance. And with the global economy still troubled, museums

must engage with the public, communicate their value and justify investment.

There are several reasons why it is hard to bridge the gulf between the realities of research and public perception of natural history. It costs money and manpower, so small, resource-strapped collections tend to have more dated displays than their wealthier counterparts. And molecular analyses, theoretical models and phylogenetic reconstructions need considerable interpretation. A computer screen full of base pairs isn't as easy to 'sell' to the public as a stuffed marmoset.

Happily, the world's biggest players are taking the problem seriously. In 2001, the American Museum of Natural History (AMNH) in New York became one of the first to tackle the molecularity of natural science with **The Genomic Revolution**, a temporary exhibition launched soon after the first draft of the human genome was published. Communicating molecular findings is hard, but it can be done.

Other developments translate more easily into displays. Take the use of computed

**NATURE.COM**  
For the evolution of  
conservation:  
[go.nature.com/tcbfk3](http://go.nature.com/tcbfk3)

tomography (CT) scanners, which expose specimens in once-impossible ways. Digital data offer a huge

opportunity for new types of presentation.

With its Darwin Centre, London's Natural History Museum has made it possible for visitors to see directly into laboratories. The 'display' has not been universally popular with researchers, but the message is clear: there are real people here, doing real science. Most museums now also give researchers and curators plenty of chances to present their work in tours, hands-on lectures and workshops.

Rethinking natural history need not mean doing away with stuffed specimens. A museum that can achieve greater parity between labs and exhibition spaces may even find it easier to embrace — and celebrate — the specimens on which it was founded. Such exhibits will be liberated, becoming objects less of natural history than of history and art. Once they acquire these new meanings, even more opportunities for public engagement will arise, and audiences will grow.

Take the Hall of North American Mammals at the AMNH, home to some of the world's most exquisite dioramas. These continue to be viable ways of conveying zoological, ecological and geological information. And the older they get, the more obvious their historical value becomes. Ongoing renovations to the gallery will see non-scientific meanings come to the fore, and will explore President Theodore Roosevelt's pioneering role in US conservation: among other things, he created five national parks and the US Forest Service.

Plenty of collections are paying more attention to the aesthetics of their specimens. The London museum was one of the first to exploit the grotesque qualities of specimens preserved in alcohol to fantastic visual effect, when it opened its spirit collection in phase 1 of the Darwin Centre in 2002. The Museum for Natural History in Berlin followed suit with its rebuilt East wing in 2010. It also boasts a stunning taxidermic 'biodiversity wall', which resembles a seventeenth-century cabinet of curiosities. And from September 2011 to January 2012, the Smithsonian Institution's National Museum for Natural History in Washington DC boasted an installation about bioluminescence by artist research fellow Shih Chieh Huang.

The nineteenth-century philosopher, poet and environmentalist Henry David Thoreau once denigrated museums as "catacombs of nature". But he never got to stand in a twenty-first-century space looking at beautifully presented specimens with a historical narrative delivered through earphones. Given that, and the ability to scan a DNA barcode to resurrect a CT-scanned, whole-organism hologram and manipulate it with a wave of his hand, Thoreau might reconsider.

Henry Nicholls is a science journalist based in London, and the author of *The Way of the Panda* (Profile Books, 2010).  
e-mail: [henry@henrynicholls.com](mailto:henry@henrynicholls.com)



# Correspondence

## Balanced regulation of synthetic biology

Contrary to the opinion of Genya Dana and colleagues (*Nature* **483**, 29; 2012), the greater sophistication of modified microbes created by synthetic biology could make them less, not more, difficult to regulate, manage and monitor than their naturally occurring counterparts.

Several routes could lead to a hypothetical disaster scenario, in which artificially produced microorganisms spread unchecked in the environment (see [go.nature.com/pncgxs](http://go.nature.com/pncgxs)). But we need to take into account that some of these paths would be favoured over others, and that ecological and other interactions might attenuate or amplify those different probabilities. Synthetic biology could then be used to engineer environmental safety into these organisms.

There is a danger that adverse public opinion could hinder the development of synthetic biology's potential. To prevent this, the media must refrain from hyping the benefits or risks of the technology. That will ensure that synthetic biology can be developed safely under a regulatory system that is based on the probability of occurrence of hazardous events, rather than on imagined possibilities.

The questions raised by Dana *et al.* should be considered as part of any risk-governance system for synthetic biology. Governance must be adaptive and scientifically and economically robust, given that most outcomes of technological innovation are still speculative. At the Economic and Social Research Council's Innogen Centre in Edinburgh, UK, we are developing such a strategy, to enable beneficial innovation with minimal risk to people and the environment while balancing stakeholder interests (see [go.nature.com/khvykc](http://go.nature.com/khvykc)).

Dana *et al.* propose that an

investment of US\$20–30 million over 10 years is needed to assess the risks associated with synthetic microbes. But, in our view, this is not yet justified. A more balanced and wide-ranging analysis is called for before major funding decisions are made relating to governance of the technology.

**Joyce Tait, David Castle** *ESRC Innogen Centre, University of Edinburgh, UK.*  
[joyce.tait@ed.ac.uk](mailto:joyce.tait@ed.ac.uk)

## Media centre: vital to US science

In my view, Colin Macilwain's reservations over the challenges facing the planned US Science Media Center are overly pessimistic (*Nature* **483**, 247; 2012).

No one — least of all the members of an exploratory committee for such a centre in the United States (including myself) — would deny that real differences exist between Britain and the United States in our cultures, politics, science and media. If established, a US centre (see [www.sciencemediacenter.org](http://www.sciencemediacenter.org)) would embrace a uniquely American model of operation to serve the country's journalists and public understanding of science. It would adapt to its cultural landscape, just as those of Canada or Japan have.

It is precisely because of the bitter contention in the United States over issues such as climate change, stem-cell research and evolution that we need a science media centre now more than ever.

Successful science media centres in other countries have demonstrated that they improve civic discourse, because journalists are better informed about the science behind the controversies. Policy-makers can make decisions based on the best available science, and citizens can make smarter personal life

choices and engage in serious political dialogue.

**Julia A. Moore** *Woodrow Wilson International Center for Scholars, Washington DC, USA.*  
[mooreblaney@gmail.com](mailto:mooreblaney@gmail.com)

## Media centre: more than public relations

As chief executive of London's Science Media Centre, I find the prospect of a similar centre in the United States more exciting than Colin Macilwain seems to (*Nature* **483**, 247; 2012). A network of science media centres (SMCs) is already flourishing in Japan, Canada, Australia, New Zealand and the United Kingdom, with Denmark soon to join the list. This model works and, given the calibre of the people involved in the US move, I am confident that they will be able to meet any challenges head-on.

The SMC model is popular with journalists because it bucks public-relations protocol. None of us promotes our brand name or any institutional message, we are just as happy to kill stories as to create them, and we embrace difficult and controversial issues. This special mix is likely to be as popular with US journalists as it is with those in the other countries.

Science journalists use quotes from SMCs and/or treat them as a source of background information. If US science journalists are less likely to use quotes, as Macilwain implies, then a US SMC would simply operate differently. However, a scarcity of specialist science journalists may force a US SMC to operate more like the Australian centre, which tends to work with general news reporters.

SMCs around the world will have to adapt to different cultures, in which attitudes to acceptable funding sources may vary. But what really matters is that SMCs should be independent of those funders, whether they be from industry, government, the media

or the scientific community.

**Fiona Fox** *Science Media Centre, London, UK.*  
[fiona@sciencemediacentre.org](mailto:fiona@sciencemediacentre.org)

## Journals' role in ethical research

False evidence in medical research can endanger lives, so countermeasures must be stepped up — particularly in light of recent cases (see [go.nature.com/9ivjnm](http://go.nature.com/9ivjnm)).

Improving scrutiny and verifying source data would be prohibitively costly and time-consuming, even if it were legally feasible. Instead, journals should demand proof before publication that research has been approved by a recognized body, such as an independent ethics committee or institutional review board, so that fraudulent claims can be identified (see, for example, *Nature* **477**, 384; 2011). This approval is legally required in most countries and has now become an international standard (see [go.nature.com/krrr6g](http://go.nature.com/krrr6g)), so is easy to obtain.

Authors and their institutions should inform the journal promptly should their results be later invalidated, and permit inspection or auditing of reports if necessary (S. Kleinert and R. Horton *Lancet* **372**, 789–790; 2008). This would speed publication of the formal retraction to correct the scientific record.

**Holger Baumgartner** *Research Ethics Committee, Innsbruck Medical University, Austria.*  
[holger.baumgartner@i-med.ac.at](mailto:holger.baumgartner@i-med.ac.at)  
*Competing financial interests declared* (see [go.nature.com/5qlre5](http://go.nature.com/5qlre5)).

### CONTRIBUTIONS

Correspondence may be sent to [correspondence@nature.com](mailto:correspondence@nature.com) after consulting the guidelines at <http://go.nature.com/cmchno>.



# David Sayre

## (1924–2012)

Crystallographer who pioneered methods of X-ray imaging and modern computing.

David Sayre, who died on 23 February, was a pioneer in crystallography and diffraction imaging, a visionary in X-ray microscopy and an architect of modern computing. A superb scientist, deep thinker and wonderful mentor, he could have built a scientific empire. But that was not his style. He was driven by the desire to do pure and original science.

Sayre was born on 2 March 1924 in New York. His father was an organic chemist whose ancestors helped to found the town of Southampton, New York, in the sixteenth century. His mother was the daughter of Jewish immigrants. Sayre was educated at Yale University in New Haven, Connecticut, graduating in 1943 at the age of 19 with a bachelor's degree in physics. The Second World War was at its height, so Sayre worked on radar at the Radiation Laboratory at the Massachusetts Institute of Technology in Cambridge.

In 1946, guessing biology would be the next exciting field, Sayre became a graduate student in biology at the University of Pennsylvania in Philadelphia and then at Harvard University in Cambridge. He was not initially interested in what he was learning, but in 1947 Sayre came across an article about X-ray crystallography that changed his life. He joined Raymond Pepinsky's crystallography laboratory at Auburn University in Alabama, where he used a mathematical operation known as the Fourier transform to analyse the structures of crystals probed with X-ray beams.

That year, Sayre married Anne Colquhoun, a fiction writer. She took a teaching position at the Tuskegee Institute, but her involvement in the school, which enrolled black students, was controversial in the Deep South at that time, and the Sayres soon left. They moved to Oxford, UK, where Sayre completed his PhD in the lab of Dorothy Hodgkin in 1951.

Sayre produced his most profound papers during this period, solving the 'phase problem' in crystallography — the loss of phase information in the measurement of diffraction intensity. In 1952, he proposed atomicity — the fact that atoms are small and discrete points relative to the space between them — as a constraint for determining the phases of crystals of small molecules, giving

rise to what is now called Sayre's equation. Atomicity is the key concept behind the direct methods used for crystallography today, although Sayre did not share the 1985 chemistry Nobel prize awarded for



it. In 1952, Sayre also realized that, even in the absence of regular crystal structure, information could be gleaned from the fine sampling of diffraction patterns.

Sayre saw early on that solving complex crystal structures would require substantial computational resources. In 1956 he joined IBM's Watson Research Center in New York, and eventually became assistant manager of the team that wrote the original FORTRAN compiler. He became corporate director of programming, and later head of the IBM programming research group. In 1969, he and his team proved the efficiency of virtual memory in computing.

In 1972–73, Sayre took a sabbatical, returning to Hodgkin's lab and to crystallography. It was during this time that one of us (J.K.) met the Sayres, forming a lasting friendship and collaboration. Anne Sayre also wrote the influential book *Rosalind*

*Franklin and DNA*, about the outstanding crystallographer and Sayre family friend who had died of cancer at an early age.

After returning to IBM, Sayre became interested in X-ray microscopy. His 1971 idea of focusing X-rays using Fresnel zone plates became a reality through the use of IBM's nanofabrication technology and with the advent of synchrotron radiation sources such as the National Synchrotron Light Source at Brookhaven National Laboratory in Upton, New York. X-ray microscopy based on zone plates is now used in synchrotron-radiation facilities worldwide.

Around 1990, Anne developed scleroderma, a debilitating disease, and David retired from work to care for her. But he continued working to realize his 1952 dream: the reconstruction of molecular structures without the use of crystals. The idea came to fruition almost 50 years later, with the publication in 1999 of the first reconstruction of a non-crystalline model object from its diffraction pattern (which was J.M.'s PhD project). This paper established coherent diffraction imaging (CDI), also called lensless imaging or diffraction microscopy, as the most promising form of high-resolution X-ray imaging. CDI is now one of the fastest-growing fields in X-ray science.

Anne died in 1998, and in the last decade of his life David suffered from Parkinson's disease. But he continued to participate in research and to offer advice. A researcher with exceptional intuition, David lived for science. His passing is a huge loss for all of us. ■

**Janos Kirz** is distinguished professor emeritus at Stony Brook University, New York, and scientific adviser for the Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. He was a collaborator and friend of David for nearly 40 years. **Jianwei Miao** is a professor in the Department of Physics and Astronomy and the California NanoSystems Institute, University of California, Los Angeles, California 90095, USA. He worked with David on coherent diffraction imaging beginning in 1996, first as a student, then as a collaborator and friend.  
e-mail: miao@physics.ucla.edu

IBM ARCHIVES

# MIDNIGHT AT THE A&E

*Be a player.*

BY TAIK HOBSON

## HAVING A BAD DAY?

Out the window from the 37th floor I fall, facing towards the heavens, arms stretched out for the divine deliverance I know will never arrive. Not a half second following my ejaculation, out hop the Minions of Bhenvul (#3: *Fear of what others might think*) in mad choreographed glee, their smiles small half moons against the starless night. I count one, two ... seven demons. Most of them will end up as demon guts on the asphalt but what difference will that make? I am one and they are minions. Stay on it! Taking aim, I activate both my RighteousRevolvers and the sky starts to rain demon blood; each shot a musical thunderclap in the key of C, true as its name[CP+5+5+5+5]. My head is still swimming from where I was struck earlier; poison of unimmortality working into my faculties, weakening my hold on the handle. And this isn't even the big boss.

## ON A GUILT TRIP?... AGAIN?

I tell myself I might still make it. Using all my Confidence Points I PURCHASE: GodHand[Grapple][CP-20/0]

## CAN'T KEEP THOSE NAGGING INSECURITIES AWAY?

GodHand is go! With seven floors to spare, I get a hold and break my fall through one of the windows, landing in a dark, empty office room sprinkled with pieces of glass. Outside I hear a cascade of demon laughter ending in a succession of splats. They think it's so funny; they don't have a clue. You try asking Emma Sophine out for a date. The thought alone invites paralysis.

I really shouldn't be thinking about this right now. Bringing up my worries is a potentially fatal slip of the mind; like a signalling beacon it attracts them. Tells them where I am.

## AT THE A&E, WE CAN HELP

Too late, I think, as the walls begin to morph. I'm out of the window just as shattered pieces of the original partition becomes a living maw, crystallized canines scraping up against my back and I know I'm in trouble. Resorting to a vertical dash, it's all I can do to concentrate on each footstep as I try to keep a better watch on my thoughts. The Null Devourer, Spawn of Nefrul (#2: *Lack of self worth*) is a pandimensional entity that becomes the antithesis of everything thrown at it. If I don't kill it now, I will be haunted by its bulbous visage for weeks to come.

Warped face in my morning cereal, corrosive odour in my sweat. Watch your dreams turn into nightmares!

Somehow, you always get more than you bargain for at the A&E. I wonder if that's why I keep coming back.

## FACE THOSE EMOTIONAL BLOCKS ONE ON ONE

I'm storming up the side of the building, dispensing Righteousness left, right, centre and just about anywhere I can, the whole structure having turned squishy and flesh-like. To hit the Devourer I'll have to move onto something smaller, but for now the strategy is to stay alive. Take that, oviposterior!



"GYA-HA-HA-HAAAAW~!"

I look up in time to catch a falling minion in the face, a late addition to the previous mess, the pair of us thrown into a complete free fall from the impact. With my revolvers lost it takes every ounce of concentration just to keep the demon's paws of wild electricity from brushing against me. In one of our twirls I register rows of lidless eyeballs where the building windows once were, moving in unison, tracking us hungrily. There's nothing sophisticated about this Devourer; it's just going for the one thing that's keeping me alive at the moment. Case in point: I blink and the building that was a writhing overfed larva a second ago is now plain bricks and mortar, whereas the line from my grapple — the GodHand that's keeping me from hitting the pavement — begins to swell and pop scales.

➤ **NATURE.COM**  
Follow Futures on  
Facebook at:  
[go.nature.com/mtoodm](http://go.nature.com/mtoodm)

This is too much.  
"Demon —" I manage under the stench

of unworldly halitosis, "— meet Null Devourer."

## SHOW THEM WHO'S IN CHARGE

Commands get scrambled thru the shock of the electrified grapple line/half-morphed Devourer, wreaking havoc on my handle and firing off stray, half-baked signals. I'm screaming. Somewhere, I can smell spit sizzling off an overheated interface ... console ... the experience is ... the experience ... what ... experience ... what am I doing ... am —

## AVATARIZE

— doing?[CP+50] —

## THEN EXORCISE

... Avatarize ... then Exorcise. The A&E ...and ... thank god for in-game commercials. My first thought when I come to is that I'm hanging by the ledge of an open window, handle realigning itself by the tips of my finger. Think. Emma Sophine. That's why you're here. So-deep-Phine-breath.

And then I see her. Ijana, Queen of Horns (#1: *Fear of rejection*). This is it. Big Boss time. The one final barrier between me and bliss. I feel my stomach seize as my handle braces for action. Deep breaths now. In less than 10 seconds she will be directly beneath me, another 20 she'll have moved into centre position, reconfiguring with her hive, making her virtually invulnerable. This is my chance. Kill her now and I can finally ask Emma out for a date. Kill her, and I will be holding That Hand.

PURCHASE: SoulSword[Blade][CP-30/20]

It's funny, the things we do to avoid facing the real thing.

I unsheath my SoulSword and jump. I have exactly no tricks left up my sleeve. The distance closes and the Queen looks up from the sound of my screaming. For a half second I forget why I do this, when I see Ijana move her own crown of horns out of my way, leaving me with a clean shot at her forehead. And the third eye embedded in there.

"DIE, DEMON! DIE!"

**THIS HAS BEEN A WORDED ACCOUNT OF AN IN-GAME SESSION. FOR OUR SPRING CAMPAIGN, PLEASE VISIT OUR WEBSITE. HAVING A BAD DAY? AT THE A&E, WE CAN HELP. ■**

*Having sworn off computer games in favour of a normal life, Taik Hobson lives in Japan.*

JACEY



## Spotlight on mobility

The complexity in patterns of human mobility, migration and communication has been difficult to unpack. Researchers have now come up with a simple theory that captures the intricacy of such phenomena. [SEE LETTER P.96](#)

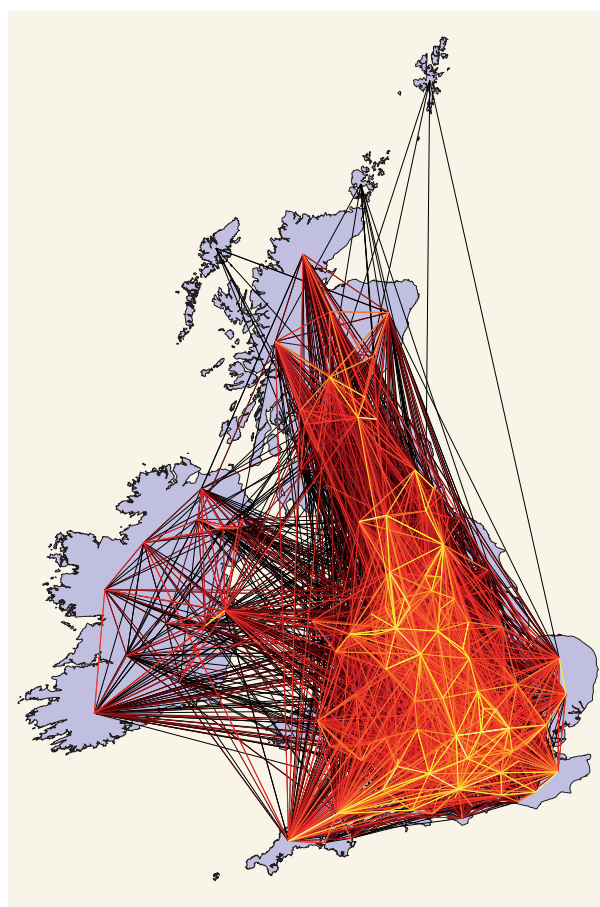
DIRK BROCKMANN

Humans are mobile, interacting animals. In this respect, we are no different from most of our fellow species: single-cell, slime-mould amoebas crawl up a chemical signal gradient to aggregate into a multicellular organism; mice seek breadcrumbs to deliver to their litters; and blue whales travel thousands of kilometres to find a mate. Most of what humans do is driven by our interactions and mobility — society doesn't work without it. Children take the bus to school and interact with their peers, employees commute to their workplaces to collaborate with their colleagues, students leave home to attend a college and subsequently seek a job in a distant city. Our movement patterns shape the structure of urban areas and the connectivity of our transportation and trade networks.

The complexity of human mobility, migration and communication patterns may suggest that no simple theory can capture these phenomena. But on page 96 of this issue, Simini *et al.*<sup>1</sup> propose a simple model that agrees with observations remarkably well.

With the emergence of pervasive data on human interactions and mobility, the past couple of years have witnessed a proliferation of empirical studies on these topics<sup>2–5</sup>. Social-network sites, location-aware devices, mobile and smart phones, e-mail communication and related technologies have generated an immense body of large-scale data sets. Numerous papers have been published reporting statistical properties, peculiarities and surprisingly simple regularities, power laws and scaling laws for human mobility. But the great mystery remains: why do we observe what we observe?

Researchers have proposed both phenomenological and ad hoc models to account for mobility data (Fig. 1). More often than not, they have had limited success. Although many of the studies are interesting, they lack the



**Figure 1 | Human-mobility pattern in the United Kingdom and Ireland.** Simini and colleagues' model<sup>1</sup> for human mobility and migration could explain complex patterns such as that seen here<sup>6</sup>. Each line represents mobility flux, the estimated amount of traffic between counties per unit time. Bright lines indicate strong flux, and darker lines denote weaker flux. This network was derived from proxy data for human mobility integrated over a period of 3 years.

heartbeat of a solid and plausible theoretical foundation that can describe observed patterns without — or, at least, with little — parameter tuning. More importantly, there are no theories that can explain the observed patterns on the basis of fundamental principles.

Lacking a sound theory, many researchers performing empirical studies on human mobility and communication have tried to fit their data to what has been termed the gravity law of transportation. This traditional

phenomenological model draws on an analogy to the force of gravity between two objects. It relies on the intuitive notion that the amount of traffic,  $T_{ij}$ , between two cities,  $i$  and  $j$ , increases with some power of their population size,  $m_i$  and  $m_j$ , and decreases with some power of the distance between them,  $d_{ij}$ , as:  $T_{ij} \approx m_i^\alpha \times m_j^\beta / d_{ij}^\gamma$ . But the model typically has three unknown parameters (the exponents  $\alpha$ ,  $\beta$  and  $\gamma$  in the equation), and almost every study that fits the gravity model to data obtains different best-fit values. Most crucially, the gravity law does not emerge from testable underlying hypotheses.

Simini and colleagues<sup>1</sup> change this situation. They propose a 'radiation model' for human mobility and migration. This model is based on two simple and plausible assumptions: humans do not enjoy moving, and they take the nearest opportunity that improves their circumstances. In other words, individuals move to a new location only because it is the closest location that offers, say, a better job. Therefore, Simini *et al.* assume that a person's geographical spotlight, centred at their current location, increases until a better place is identified, but no farther. The crucial assumption in the radiation model is that individuals do not necessarily seek the best opportunity, but rather their priority is to pick the closest destination.

By assuming that the number of opportunities offered at a location is proportional to the population size of that location, and that each opportunity has a random quality score, the researchers<sup>1</sup> were able to compute the expected flux of individuals between two locations, making the spatial distribution of the population the only input to their theory. The authors then compared the theory's predictions with multiple data sets ranging from daily-commute traffic to long-term migration patterns, mobile-phone mobility and communication patterns. They find that this theory substantially outperforms the gravity model.

D. BROCKMANN



Interestingly, their theory reduces to a specific type of gravity model when a homogeneous distribution of the populations (and thus opportunities) is assumed. This could explain why the gravity-theory analogy has been so persistent in mobility research, and also indicates that it is equivalent to an idealized scenario, one hardly ever encountered in real-life settings.

The radiation model may provide a route to further exploration. It could be useful for researchers interested in understanding processes mediated by human mobility, such as the introduction of animals and plants into a

new habitat and the spread of human infectious diseases<sup>2</sup>. Until now, computational models in these areas relied on direct data implementation and gravity models to fill gaps in incomplete mobility data sets.

A pertinent question for future work is why some societies are more mobile than others. A comparative analysis between the United States and European countries would be a promising starting point for addressing this issue. ■

**Dirk Brockmann** is at the Northwestern Institute on Complex Systems, and in

the Department of Engineering Sciences and Applied Mathematics, Northwestern University, Evanston, Illinois 60208-3125, USA.

e-mail: [brockmann@northwestern.edu](mailto:brockmann@northwestern.edu)

1. Simini, F., González, M. C., Maritan, A. & Barabási, A.-L. *Nature* **484**, 96–100 (2012).
2. Brockmann, D., Hufnagel, L. & Geisel, T. *Nature* **439**, 462–465 (2006).
3. González, M. C., Hidalgo, C. A. & Barabási, A.-L. *Nature* **453**, 779–782 (2008).
4. Vespignani, A. *Science* **325**, 425–428 (2009).
5. Brockmann, D. *Phys. World* Feb., 31–34 (2010).
6. Brockmann, D. & Theis, F. *IEEE Pervasive Comput.* **7**(4), 28–35 (2008).

## CLIMATE CHANGE

# A tale of two hemispheres

**A reconstruction of temperature from proxy records shows that the rise in global mean temperature closely resembled, but slightly lagged, the rise in carbon dioxide concentration during the last period of deglaciation. [SEE ARTICLE P.49](#)**

ERIC W. WOLFF

Between about 19,000 and 10,000 years ago, Earth emerged from the last glacial period. The whole globe warmed, ice sheets retreated from Northern Hemisphere continents and atmospheric composition changed significantly. Many theories try to explain what triggered and sustained this transformation (known as the glacial termination), but crucial evidence to validate them is lacking. On page 49 of this issue, Shakun *et al.*<sup>1</sup> use a global reconstruction of temperature to show that the transition from the glacial period to the current interglacial consisted of an antiphased temperature response of Earth's two hemispheres, superimposed on a globally coherent warming. Ocean-circulation changes, controlling the contrasting response in each hemisphere, seem to have been crucial to the glacial termination.

The centrepiece of Shakun and colleagues' study<sup>1</sup> is a reconstruction — using 80 marine, terrestrial and ice-core proxy records — of the latitudinal temperature pattern and global mean temperature throughout the termination. This is a major achievement: the difficulties of synchronizing the records and of ensuring that they are sufficiently representative of the whole planet, are considerable. Global mean temperature rose in two main steps, closely mirroring the rise in atmospheric carbon dioxide measured<sup>2</sup> in Antarctic ice cores (see Fig. 2 of the paper<sup>1</sup>). And in contrast to Antarctic temperature<sup>2</sup>, global mean temperature lagged carbon dioxide rise

by  $460 \pm 340$  years during the termination.

As anticipated from comparisons of Antarctic with Greenland temperatures<sup>3</sup>, the Northern and Southern Hemispheres show different patterns of temperature change (see Fig. 4 of the paper<sup>1</sup>). The authors<sup>1</sup> quantify this difference by subtracting the mean temperature of the Southern Hemisphere from that of the Northern. This gives a W-shaped profile that is remarkably similar in timing and shape to a geochemical record<sup>4</sup> that is considered to be a proxy for the strength of the Atlantic meridional overturning circulation (AMOC). The AMOC is the branch of ocean circulation in the Atlantic that takes warm surface waters northwards, balanced by the flow of cold deep water southwards. The strength of this circulation has a considerable impact on the transfer of heat between the hemispheres.

Some studies have proposed<sup>5,6</sup> that changes in ocean heat transport are an essential part of glacial termination. Shakun *et al.*<sup>1</sup> combine their data with simulations based on an ocean–atmosphere general circulation model to present a plausible sequence of events from about 19,000 years ago onwards. They propose that a reduction in the AMOC (induced in the model by introducing fresh water into the North Atlantic) led to Southern Hemisphere warming, and a net cooling in the Northern Hemisphere. Carbon dioxide concentration began to rise soon afterwards, probably owing to degassing from the deep Southern Ocean; although quite well documented, the exact combination of mechanisms for this rise remains a subject of debate. Both hemispheres

then warmed together, largely in response to the rise in carbon dioxide, but with further oscillations in the hemispheric contrast as the strength of the AMOC varied. The model reproduces well both the magnitude and the pattern of global and hemispheric change, with carbon dioxide and changing AMOC as crucial components.

The success of the model used by Shakun and colleagues in reproducing the data is encouraging. But one caveat is that the magnitude of fresh water injected into the Atlantic Ocean in the model was tuned to produce the inferred strength of the AMOC and the magnitude of interhemispheric climate response; the result does not imply that the ocean circulation in the model has the correct sensitivity to the volume of freshwater input<sup>7</sup>.

Shakun and colleagues' work does provide a firm data-driven basis for a plausible chain of events for most of the last termination. But what drove the reduction in the AMOC 19,000 years ago? The authors<sup>1</sup> point out that there was a significant rise in temperature between 21,500 and 19,000 years ago in the northernmost latitude band (60–90°N). They propose that this may have resulted from a rise in summer insolation (incoming solar energy) at high northern latitudes, driven by well-known cycles in Earth's orbit around the Sun. They argue that this rise could have caused an initial ice-sheet melt that drove the subsequent reduction in the AMOC.

However, this proposal needs to be treated with caution. First, there are few temperature records in this latitude band: the warming is seen clearly only in Greenland ice cores. Second, there is at least one comparable rise in temperature in the Greenland records, between about 62,000 and 60,000 years ago, which did not result in a termination. Finally, although it is true that northern summer insolation increased from 21,500 to 19,000 years ago, its absolute magnitude remained lower than at any time between 65,000 and 30,000 years ago. It is not clear why an increase in insolation from a low value initiated termination whereas a continuous period of higher insolation did not.

In short, another ingredient is needed to

explain the link between insolation and termination, and the triggers for the series of events described so well in Shakun and colleagues' paper<sup>1</sup>. The see-saw of temperature between north and south throughout the glacial period, most clearly observed in rapid Greenland warmings (Dansgaard–Oeschger events), is often taken as a sign that numerous changes in AMOC strength occurred. However, the AMOC weakening that started 19,000 years ago lasted for much longer than previous ones, allowing a much more substantial rise in southern temperature and in carbon dioxide concentration. Why was it so hard, at that

time, to reinvigorate the AMOC and end this weakening? And what is the missing ingredient that turned the rise in northern insolation around 20,000 years ago into the starting gun for deglaciation, when higher insolation at earlier times failed to do so? It has been proposed<sup>8</sup> that terminations occur only when northern ice-sheet extent is particularly large. If this is indeed the extra ingredient, then the next step in unwinding the causal chain must be to understand what aspect of a large ice sheet controls the onset and persistence of changes in the AMOC that seem to have been key to the last deglaciation. ■

**Eric W. Wolff** is at the British Antarctic Survey (Natural Environment Research Council), High Cross, Cambridge CB3 0ET, UK. e-mail: ewwo@bas.ac.uk

1. Shakun, J. D. *et al.* *Nature* **484**, 49–54 (2012).
2. Monnin, E. *et al.* *Science* **291**, 112–114 (2001).
3. EPICA Community Members *Nature* **444**, 195–198 (2006).
4. McManus, J. F. *et al.* *Nature* **428**, 834–837 (2004).
5. Denton, G. H. *et al.* *Science* **328**, 1652–1656 (2010).
6. Wolff, E. W., Fischer, H. & Röthlisberger, R. *Nature Geosci.* **2**, 206–209 (2009).
7. Valdes, P. *Nature Geosci.* **4**, 414–416 (2011).
8. Parrenin, F. & Paillard, D. *Earth Planet. Sci. Lett.* **214**, 243–250 (2003).

## NEUROSCIENCE

# The symphony of choice

**The brain's parietal cortex seems to orchestrate decision-making without single neurons performing 'solos'. Rather, decision-specific motifs emerge as highly organized sequences of short-lived neuronal activity. [SEE ARTICLE P.62](#)**

EDUARDO DIAS-FERREIRA & RUI M. COSTA

Sometimes it may be necessary to turn down that Vivaldi concerto while driving, to catch up with the traffic news. But a warning of an upcoming traffic jam would be totally useless if the driver could not sort out the specific sequences of left and right turns along an alternative itinerary. Such transitions from perceptual decisions (choosing one option on the basis of sensory information) to movement planning rely on the brain's parietal cortex<sup>1,2</sup>. But what neuronal dynamics underlie such complex chains of events? On page 62 of this issue, Harvey and colleagues<sup>3</sup> report decision-specific sequences of neuronal activity in the parietal cortex of mice during a memory-guided navigation task. Remarkably, the animals' choices were not signalled by sustained activity of particular neurons during the decision-making period, as has been shown in other studies<sup>1,2</sup>. Rather, the choice was indicated by the activation of specific sequences of intermingled neuronal populations, with each neuron transiently active during a particular period of the task.

Harvey *et al.*<sup>3</sup> used a virtual T-shaped maze that mice, with their heads restrained, explored by running on a spherical treadmill. At the same time, the authors tracked the activity of individual neurons in a specific layer of the animals' parietal cortex by imaging cellular calcium levels — an indication of neuronal activity — using two-photon microscopy. In each trial, the mouse would actively navigate through a first part of the maze, where a visual

cue would be presented. After that, there was a delay period in which it would continue straight ahead until it faced a crossroad where, depending on the initial cue, it had to take either a left or a right turn to reach a reward (Fig. 1a).

The authors recorded the activity of sufficient numbers of neurons simultaneously to notice that neurons with sustained activity during an entire task period (cue, delay or decision) were rare. Moreover, the choice made by the mouse — left or right turn — could be predicted at any point in the maze from a sparse pattern of the neuronal activity at the ensemble level. From a bird's-eye view, these activation patterns showed an ordered progression through specific neuronal populations, reflecting not merely the spatial and temporal progression of the mouse through the maze, but its future choice (Fig. 1b).

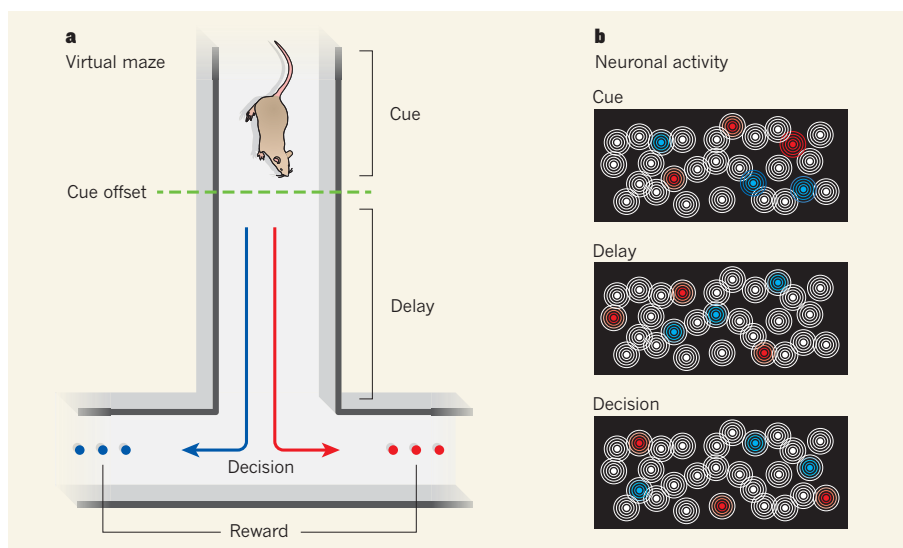
Compared with other studies<sup>1,2</sup>, neurons with sustained activity were rare in these experiments<sup>3</sup>. This disparity probably stems from differences in task structure between the studies. In Harvey and colleagues' virtual maze<sup>3</sup>, the mice were allowed to continue moving during the delay period, whereas in the other studies<sup>1,2</sup> they were required to stay immobile during that period, fixating on a point on a screen. Hence, one intriguing possibility would be that the sustained activity of certain neurons observed by other authors during the delay period represents a slowing down of ensemble dynamics during 'fixation'. It is interesting that — in the present study<sup>3</sup>, as in previous reports<sup>1,2</sup> — the divergence of neuronal dynamics for different choices starts

from the moment of cue presentation, before the delay period.

Could the neuronal dynamics reported by Harvey and colleagues<sup>3</sup> be related to visual information processing or motor performance, rather than to choice or action planning? This does not seem to be the case, as there was no correlation between the animals' running trajectories and the neuronal activity before choice. Furthermore, when the authors injected a drug to inactivate the parietal cortex, the animal's performance of the (memory-guided) task was impaired. But this was not the case for a visually guided version of the task, in which the visual cue was presented at the crossroad, rather than earlier in the maze. Moreover, during virtual linear-track tests (which required running and turning behaviours similar to those in the T-maze), and during open-loop experiments (in which images simulating runs through the maze were played to the mice), the parietal cortex was less engaged and less tuned to specific maze locations than during the memory-guided T-maze experiment. Therefore, the choice-specific sequential activation of neurons along the maze probably reflects the emergence of functional motifs for action planning.

Since Hubel and Wiesel's discovery<sup>4</sup> in the 1950s that neurons are arranged in functional columns in the brain's visual cortex, it has been tempting to think of cortical circuits as being organized in anatomically discrete functional domains. Harvey *et al.*<sup>3</sup> do find that the activities of the parietal-cortex neurons are slightly more closely correlated with those of nearer neurons than with those of more distant ones. However, choice-specific functional motifs are intermingled: there was no apparent local organization of neurons tuned for specific task periods or for specific choice preferences (Fig. 1b). It is not clear whether the type of spatial organization reported by the authors is an exception to Hubel and Wiesel's initial observations<sup>4</sup> or a more common mode of organization than previously thought, given the growing body of evidence for anatomically intermixed functional motifs<sup>5,6</sup>, especially in mouse brain circuits.





**Figure 1 | Neuronal ensembles for decision-making.** Memory-guided choices rely on activity in the brain's parietal cortex. **a**, Harvey *et al.*<sup>3</sup> set up a virtual T-maze in which mice had to choose either a left or a right turn to reach a reward, based on a previously displayed visual cue. **b**, The authors monitored neuronal activity in the mouse parietal cortex while the animals performed the task. An idealized microscopic view of the neurons is shown, in which neurons are coloured either blue or red to indicate activity when the mice took the correct turn (left or right, depending on the cue). The researchers found that the parietal cortex shows transient neuronal activity during specific periods of the task (cue, delay and decision), and that the activities of individual neurons form sequences that are choice-specific for the correctly performed trials. Although highly organized, these neuron-population dynamics emerge without apparent anatomical organization.

It has been found<sup>7</sup> that intermingled circuits can implement functional connectivity. In fact, Harvey *et al.*<sup>3</sup> provide indirect evidence that supports the emergence of functionally connected neuronal assemblies, much like those proposed by Hebb<sup>8</sup> in the 1940s, given that neurons that are sequentially activated during correctly performed trials have similar activity relationships during incorrectly performed trials and inter-trial intervals. This observation raises the possibility that — as has been observed in other neuronal circuits<sup>9,10</sup> — the sequential activation of neurons in the parietal cortex emerges from internal dynamics, probably reflecting the connectivity patterns in the circuit.

Accumulated evidence from studies in various animals — from leeches<sup>11</sup> to songbirds<sup>12</sup> and rodents<sup>3,9</sup> — points towards sequence-based dynamics as a common mechanism underlying action planning. The ordered progression of neuronal activity through a population of neurons suggests that indexing of information or time coordination occurs as an action is being performed. So, in real-world situations, competition between different actions might be established as competition between different paths of activity at the neural-circuit level. According to this view, particular environmental features would trigger sequences of neuronal activity that would lead to specific actions depending on the functional connectivity between sequence elements. Similarly, the presence of common elements in different sequences of neuronal activity could

generate switching points ('Am I right, or am I wrong?' moments), in which the behavioural output would depend on the internal properties of the circuit and the weight of each environmental feature at that particular moment.

#### ATOMIC PHYSICS

## An almost lightless laser

**Lasers are often described in terms of a light field circulating in an optical resonator system. Now a laser has been demonstrated in which the field resides primarily in the atomic medium that is used to generate the light. SEE LETTER P.78**

VLADAN VULETIC

**A**tomic clocks based on optical transitions in trapped ions or atoms are the most accurate instruments ever made<sup>1–4</sup>. However, the best atomic clocks are limited by the frequency stability of the laser with which the optical transitions are probed. In a study that builds on previous theoretical research<sup>5</sup>, Bohnet *et al.*<sup>6</sup> (page 78) describe a prototype of an unusual ultrastable laser.

Atomic transitions used for optical clocks have spectral linewidths of the order of millihertz, and standard 'free-running' lasers are not sufficiently stable in frequency to probe such ultra-narrow transitions directly. Therefore, the best atomic clocks rely on lasers

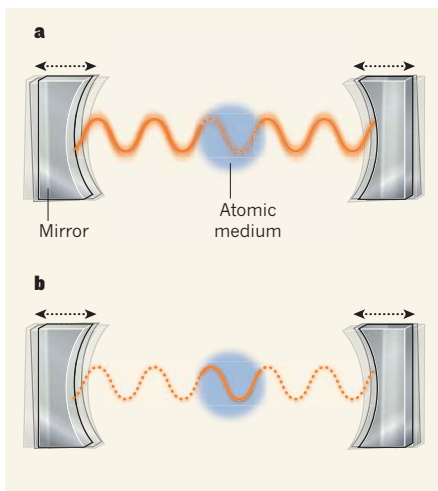
The parietal cortex is interconnected with other brain areas involved in, for example, sensory and motor processing, memory and decision-making. Such a strategic position, together with the neuronal-population dynamics now reported by Harvey *et al.*<sup>3</sup>, places this brain region at the heart of the above-mentioned competition between different actions. Researchers will probably not extract a decision-making algorithm solely from decoding the parietal-cortex circuitry, but understanding the parietal orchestra will certainly shine a light on the principles of action planning. ■

**Eduardo Dias-Ferreira and Rui M. Costa** are at the Champalimaud Neuroscience Programme, Champalimaud Center for the Unknown, 1400-038 Lisbon, Portugal. e-mail: ruicosta@fchampalimaud.org

- Shadlen, M. N. & Newsome, W. T. *J. Neurophysiol.* **86**, 1916–1936 (2001).
- Andersen, R. A. & Buneo, C. A. *Annu. Rev. Neurosci.* **25**, 189–220 (2002).
- Harvey, C. D., Coen, P. & Tank, D. W. *Nature* **484**, 62–68 (2012).
- Hubel, D. H. & Wiesel, T. N. *J. Physiol. (Lond.)* **148**, 574–591 (1959).
- Ohki, K., Chung, S., Ch'ng, Y. H., Kara, P. & Reid, R. C. *Nature* **433**, 597–603 (2005).
- Komiyama, T. *et al. Nature* **464**, 1182–1186 (2010).
- Ko, H. *et al. Nature* **473**, 87–91 (2011).
- Hebb, D. O. *The Organization of Behavior: A Neuropsychological Theory* (Wiley, 1949).
- Pastalkova, E., Itskov, V., Amarasingham, A. & Buzsaki, G. *Science* **321**, 1322–1327 (2008).
- Dragoi, G. & Tonegawa, S. *Nature* **469**, 397–401 (2011).
- Briggman, K. L., Abarbanel, H. D. & Kristan, W. B. Jr *Science* **307**, 896–901 (2005).
- Hahnloser, R. H., Kozhevnikov, A. A. & Fee, M. S. *Nature* **419**, 65–70 (2002).

whose frequency is stabilized using an external reference optical resonator — an arrangement of two highly reflective mirrors that allows light to bounce back and forth between them many times.

Currently, such reference resonators achieve a fractional frequency stability of one part in  $10^{15}$ . This value corresponds to a change in the resonator's length (the distance between the mirrors), caused by vibrations in the mirrors, of less than the radius of a proton. Below this level, the stability of the reference resonator is limited by thermal noise in the mirrors, a fundamental process that leads to fluctuations in the resonator's effective length, and thus in the frequency of the laser. Although further progress is difficult on this front, it may be possible



**Figure 1 | Standard and superradiant lasers.** **a**, In a standard laser, the amplitude and phase of the laser light are mostly stored in the light field that circulates between two reflective mirrors, rather than in the atomic medium that is used to generate the laser light. Vibrations in the mirrors, and so in the distance between them, lead to variations in the laser's frequency. **b**, In a superradiant laser such as that demonstrated by Bohnet *et al.*<sup>6</sup>, the amplitude and phase are mostly stored in the atomic medium, and the laser frequency depends only very weakly on the distance between the mirrors.

to increase the frequency stability by operating the system at cryogenic temperatures and using mirror materials that have improved mechanical properties.

To realize an ultrastable laser that does not require an external reference resonator, theorists have proposed<sup>5</sup> a laser that operates in an unusual regime, in which the spectral linewidth of the atomic medium used to generate the laser light is much smaller than the linewidth of the optical resonator in which the medium is placed. In such a system, which may be termed a 'superradiant' laser after early work by the physicist Robert Dicke<sup>7</sup>, the laser's energy is stored predominantly inside the atoms, rather than in the light field circulating inside the resonator (Fig. 1). This makes the laser frequency largely immune to changes in the resonator's length, with the extent of immunity given by the ratio of the linewidth of the atomic medium to that of the resonator.

Bohnet and colleagues<sup>6</sup> used an atomic medium consisting of rubidium atoms, a species that is easy to cool and trap but for which no narrow-linewidth atomic transition is readily available for lasing. To produce a superradiant laser, the authors resort to a neat trick: a narrow atomic transition can be mimicked by employing an external laser to weakly drive a transition between two long-lived atomic ground states. In this case, the emitted laser light is not absolutely stable in frequency, but only when measured relative to the frequency of the driving laser. Nevertheless, the authors were able to use their set-up to

test key predictions<sup>5</sup> for the superradiant laser.

Although similar lasing has been observed before in a cold-atom system<sup>8,9</sup>, Bohnet *et al.* are the first to characterize the frequency stability of the laser, and to demonstrate explicitly that the frequency of the laser depends only very weakly on the resonator's length. The authors find that the effect of changes in the resonator's length on the frequency of the laser is 10,000 times less than that observed for a standard laser.

Remarkably, the laser can be operated when the resonator contains, on average, less than one photon. Furthermore, the authors show that if the light field inside the resonator is completely turned off by switching off the driving laser, the phase of the laser light (the timing of the electromagnetic wave's peaks and troughs) can be preserved in the atoms for several milliseconds before the driving laser is turned back on. This observation demonstrates that the laser's electromagnetic waves are stored inside the atomic ensemble, as first predicted by Dicke<sup>7</sup>.

Although further tests of the frequency stability need to be performed with a system that operates on an absolutely narrow and

stable atomic-clock transition, Bohnet and colleagues' work<sup>6</sup> is encouraging, and paves the way towards a scheme with the potential to significantly improve the precision of atomic clocks. Ever more accurate timekeeping not only has a variety of technological applications, such as in telecommunication networks and the Global Positioning System, but will also allow unprecedented tests of some of the basic laws that govern our Universe. ■

**Vladan Vuletic** is in the Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.  
e-mail: vuletic@mit.edu

1. Rosenband, T. *et al. Science* **319**, 1808–1812 (2008).
2. Ludlow, A. D. *et al. Science* **319**, 1805–1808 (2008).
3. Jiang, Y. Y. *et al. Nature Photon.* **5**, 158–161 (2011).
4. Katori, H. *Nature Photon.* **5**, 203–210 (2011).
5. Meiser, D., Ye, J., Carlson, D. R. & Holland, M. J. *Phys. Rev. Lett.* **102**, 163601–163604 (2009).
6. Bohnet, J. *et al. Nature* **484**, 78–81 (2012).
7. Dicke, R. H. *Phys. Rev.* **93**, 99–110 (1954).
8. Hilico, L., Fabre, C. & Giacobino, E. *Europhys. Lett.* **18**, 685–688 (1992).
9. Chan, H. W., Black, A. T. & Vuletic, V. *Phys. Rev. Lett.* **90**, 063003–063006 (2003).

#### CANCER

## Limitations of therapies exposed

**Certain drugs that are used to treat cancer affect blood-vessel formation in tumours. But it seems that these antiangiogenic drugs can reduce the efficiency of other anticancer agents and increase the tumours' aggressiveness.**

**ORIOL CASANOVAS**

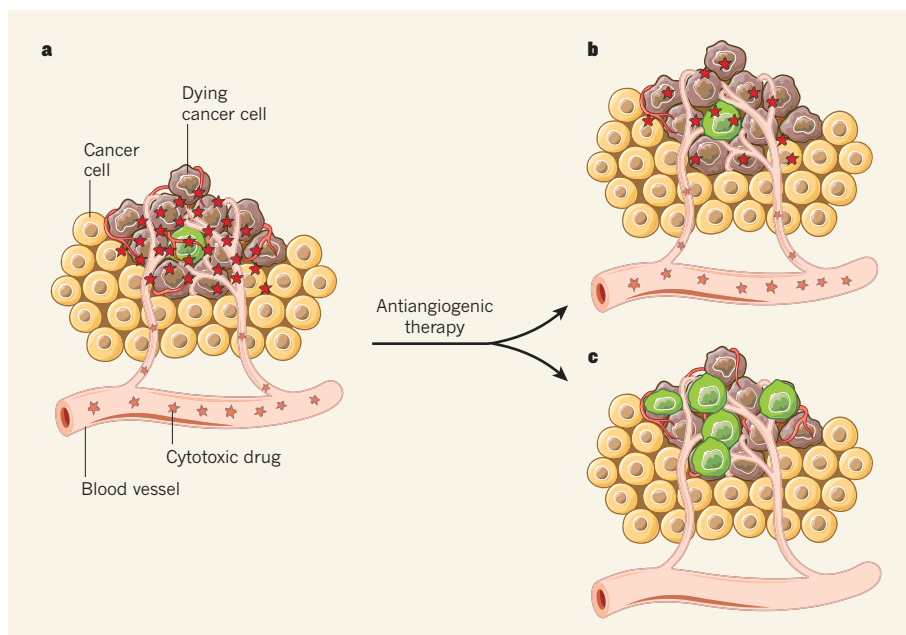
**T**umour growth depends on angiogenesis, the formation of new blood vessels, to ensure a continuous supply of oxygen and nutrients. That is why anti-angiogenic agents are used to treat certain cancers, either alone or in combination with traditional cytotoxic drugs. However, the mechanistic details of how these combination therapies work are far from clear, and accumulating evidence is exposing their limitations. Writing in *Cancer Cell*, Van der Veldt *et al.*<sup>1</sup> report that angiogenesis inhibitors can decrease the delivery of cytotoxic drugs to tumours in patients, and hence hinder the drugs' therapeutic benefits. And in a paper published in *Proceedings of the National Academy of Sciences*, Conley *et al.*<sup>2</sup> find that tumours can adapt to antiangiogenic therapy by accumulating particularly aggressive cells.

The main target of current antiangiogenic agents is a protein called vascular endothelial

growth factor (VEGF), which has a central role in angiogenesis. Although it has been known for several years that VEGF inhibitors provide additional antitumour effects when combined with cytotoxic drugs (Fig. 1a), the underlying mechanism has been a mystery since the early positive results of drug-combination trials<sup>3</sup>. The most widespread explanation for such a mechanism, the 'vascular normalization' theory, was proposed in 2001. According to this theory<sup>4,5</sup>, antiangiogenic therapy induces structural and functional changes in tumour blood vessels — which have abnormal features — to make them more similar to normal vessels and, as a result, blood flow is increased and cytotoxic drugs can more easily enter the tumour.

To test the theory in a clinical setting, Van der Veldt and colleagues<sup>1</sup> studied the uptake and retention of a cytotoxic drug (docetaxel) in 10 patients with advanced-stage non-small-cell lung cancer (NSCLC). By using radiolabelled docetaxel together with a





**Figure 1 | Drawbacks of anticancer therapies.** **a**, Certain cancers are treated with antiangiogenic drugs (which affect blood-vessel formation), either alone or in combination with cytotoxic agents that inhibit the growth of cancer cells. Some of these cells (green) are particularly dangerous because they can be more resistant to cytotoxic therapy than the other tumour cells and can spread to other organs to seed new tumours. **b,c**, A decreased blood supply to the tumour, which is the main benefit from antiangiogenic therapy, is also the basis for the therapy's limitations. **b**, It can reduce the distribution of cytotoxic agents in the tumour, and hence their efficacy. Van der Veldt *et al.*<sup>1</sup> report one such undesirable outcome in patients with non-small-cell lung cancer. **c**, By reducing oxygen levels in the tumour, antiangiogenic drugs can induce the accumulation of more aggressive cells that have an increased capacity to spread to other organs. Conley *et al.*<sup>2</sup> document this phenomenon in mouse models of breast cancer in the absence of a cytotoxic agent.

sensitive imaging method (positron emission tomography), the authors demonstrate that VEGF inhibition with a drug called bevacizumab induces a fast and sustained decrease — not increase — in the penetration of both water and docetaxel in the tumours (Fig. 1b). These results contrast with those of previous studies in patients with rectal cancer<sup>6</sup> and in patients with glioblastoma (a brain tumour)<sup>7</sup>, which showed that bevacizumab treatment induces vascular normalization and increased glucose uptake in the tumours. However, glucose uptake by tumour cells does not, in my opinion, necessarily correlate with cytotoxic-drug delivery and penetration into tumours.

The discrepancies between the authors' observations<sup>1</sup> and previous results<sup>6,7</sup> could also be due to differences in blood-vessel networks and in the response to angiogenesis inhibitors between the three cancer types, as it is known that these agents can affect blood vessels in different ways in different tissues<sup>7,8</sup>. In any case, the finding<sup>1</sup> that, at least in patients with NSCLC, antiangiogenic therapy does not improve cytotoxic drug delivery to tumours — but rather has the opposite effect — exposes a perturbing drawback to such treatments. Indeed, this could be the cause of the modest benefits of these combination therapies in NSCLC and other tumour types<sup>9</sup>. Such a potential shortcoming could be circumvented

by optimizing the scheduling of the therapeutic agents. For example, rather than administering both types of drug to a patient during the same period, treatment with blood-distributed cytotoxic agents could be followed by antiangiogenic therapy.

Other limitations of the use of angiogenesis inhibitors derive from the fact that tumours are highly adaptable. Although their ability to become resistant to cytotoxic drugs and radiation — another common anticancer therapy — has long been known, it was initially postulated<sup>10</sup> that antiangiogenic drugs would not suffer from the same problem because they target blood vessels rather than tumour cells. Yet preclinical and clinical evidence<sup>11,12</sup> has revealed that tumours can indeed adapt and become resistant to antiangiogenic therapy.

As if that was not bad enough, angiogenesis inhibitors have been shown<sup>13,14</sup> to make some tumours more aggressive in animal models (Fig. 1c). To explore this issue, Conley and co-workers<sup>2</sup> implanted human cancer cells (derived from established breast-cancer cell lines) in mice. They then treated the animals with the antiangiogenic agents sunitinib — which inhibits VEGF's main cell-surface receptors — and bevacizumab. The treatment induced an accumulation of certain cancer cells that expressed the enzyme aldehyde dehydrogenase and that, like cancer

progenitor cells, could initiate tumours when reimplanted in other mice. Similar cell populations have been described in tissue samples from patients with inflammatory breast cancer<sup>15</sup> and in glioblastoma in mice given combination therapies<sup>16</sup>.

Conley *et al.*<sup>2</sup> go on to delineate a possible cellular and molecular mechanism for the increased aggressiveness and spread capacity of tumours treated with antiangiogenic drugs. They find that the drugs, by inducing oxygen deficiency (hypoxia) in the tumours, activate not only a hypoxia-response program but also the Akt/ $\beta$ -catenin signalling pathway, which regulates cell growth and adhesion between cells. This pathway has been previously implicated in the regulation of breast-cancer progenitor cells<sup>17</sup>. The authors suggest that the drug-induced hypoxia response activates the Akt/ $\beta$ -catenin pathway, which in turn stimulates the growth of specific, more aggressive, cancer-cell populations.

How could this drawback of antiangiogenic therapies be overcome? One possibility would be to combine angiogenesis inhibitors with drugs that suppress the cancer cells' response to hypoxia, or with inhibitors of the Akt/ $\beta$ -catenin pathway. Another alternative could be the use of molecules such as modified semaphorin proteins<sup>18</sup>, which can exert dual (or multiple) anticancer effects by simultaneously targeting angiogenesis and blocking tumour spread.

Overall, the papers by Van der Veldt *et al.*<sup>1</sup> and Conley *et al.*<sup>2</sup> emphasize the need for a carefully balanced evaluation of the benefits and limitations of antiangiogenic therapies. As mentioned above, such treatments could be improved by sequential scheduling of cytotoxic and antiangiogenic drugs, or by smarter combinations of these drugs with others targeting progenitor-cell pathways. In any case, despite many open questions, there is hope that an understanding of the therapies' weaknesses will translate into therapeutic gains. ■

**Oriol Casanovas** is in the Translational Research Laboratory, Catalan Institute of Oncology, Bellvitge Institute for Biomedical Research, E-08907 L'Hospitalet de Llobregat, Spain.  
e-mail: ocasanovas@iconcologia.net

1. Van der Veldt, A. A. M. *et al.* *Cancer Cell* **21**, 82–91 (2012).
2. Conley, S. J. *et al.* *Proc. Natl Acad. Sci. USA* **109**, 2784–2789 (2012).
3. Hurwitz, H. I. *et al.* *J. Clin. Oncol.* **23**, 3502–3508 (2005).
4. Jain, R. K. *Nature Med.* **7**, 987–989 (2001).
5. Jain, R. K. *Science* **307**, 58–62 (2005).
6. Willett, C. G. *et al.* *Nature Med.* **10**, 145–147 (2004).
7. Batchelor, T. T. *et al.* *Cancer Cell* **11**, 83–95 (2007).
8. Kamba, T. *et al.* *Am. J. Physiol. Heart Circ. Physiol.* **290**, H560–H576 (2006).
9. Sandler, A. *et al.* *N. Engl. J. Med.* **355**, 2542–2550 (2006).
10. Boehm, T., Folkman, J., Browder, T. & O'Reilly, M. S. *Nature* **390**, 404–407 (1997).
11. Bergers, G. & Hanahan, D. *Nature Rev. Cancer* **8**,

- 592–603 (2008).  
 12. Rini, B. I. & Atkins, M. B. *Lancet Oncol.* **10**, 992–1000 (2009).  
 13. Páez-Ribes, M. *et al. Cancer Cell* **15**, 220–231 (2009).  
 14. Ebos, J. M. L. *et al. Cancer Cell* **15**, 232–239 (2009).

15. Charafe-Jauffret, E. *et al. Clin. Cancer Res.* **16**, 45–55 (2010).  
 16. Folkins, C. *et al. Cancer Res.* **67**, 3560–3564 (2007).  
 17. Korkaya, H. *et al. PLoS Biol.* **7**, e1000121 (2009).  
 18. Casazza, A. *et al. EMBO Mol. Med.* **4**, 234–250 (2012).

## GENOMICS

# Stickleback is the catch of the day

Whole-genome sequences from a marine fish that has adjusted to life in fresh water give hints about general genetic mechanisms that drive the evolution of adaptations to new environmental niches. [SEE ARTICLE P.55](#)

HOPI E. HOEKSTRA

The traditional recipe for publication of a genome sequence goes something like this: one part ‘biology’ (an often flowery description of the distinctive aspects of the organism whose genome has been sequenced); two parts ‘assembly and annotation’ (how the latest DNA sequencing and computational technologies were used to produce a high-quality sequence); and three parts ‘comparative analyses’ (for example, observations of rapidly evolving genes, or expansion or loss of gene families). All this is followed by a dash of ‘fun speculation’ on how these distinctive genomic characteristics might yield insight into the biology of the creature under study.

Although each genome paper has its own twist, such as filling in a critical gap in the tree of life<sup>1</sup>, a novel sequencing and assembly strategy<sup>2</sup>, or proof that an entire genome can be sequenced by a single laboratory<sup>3</sup>, the

main contribution of such work is usually the sequence itself — as a tool to delve deeper into biological questions. But there are new cooks in the kitchen, and some of them are following a different recipe. On page 55 of this issue, Jones *et al.*<sup>4</sup> report the first whole-genome sequence of the threespine stickleback fish, and at the same time reel in answers to some big questions in evolutionary biology.

The ancestral form of the threespine stickleback (*Gasterosteus aculeatus*) was a marine fish with protective bony ‘armour’ that repeatedly colonized lakes and streams around the world following the retreat of glaciers at the end of the last ice age some 10,000 years ago. In their new freshwater habitats, these fish repeatedly evolved new morphological and physiological adaptations, such as loss of the bony plating and spines and a change in salinity tolerance, and this evolutionary pattern has elevated the threespine stickleback to ‘supermodel’ status<sup>5</sup> in studies of adaptation and speciation. The

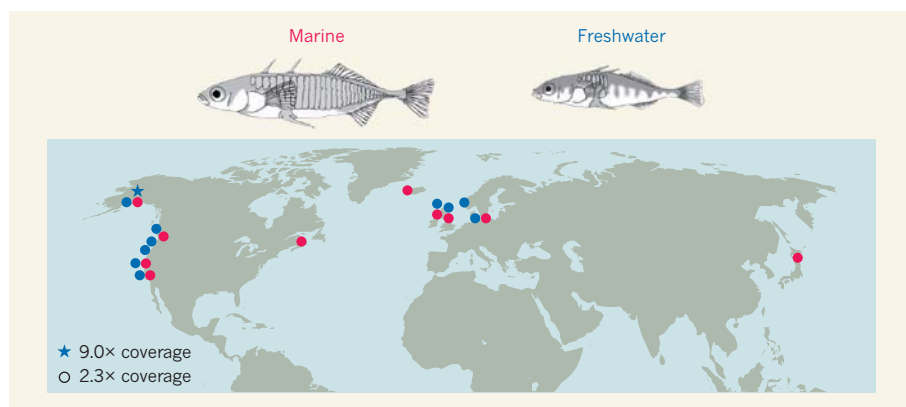
one missing piece has been the complete genome sequence. Jones and colleagues<sup>4</sup> provide just this, with a high-quality sequence of a freshwater Alaskan stickleback.

But the authors don’t stop there. They also report sequence information from another 20 sticklebacks sampled around the globe (Fig. 1) that form 10 geographically linked pairs of marine and freshwater fish. And they develop two complementary approaches — one based on phylogenetic relationships and a second on pairwise genetic distance — to scan the genomes for regions that are substantially different between marine and freshwater sticklebacks but similar within marine, or freshwater, populations from different geographical locations. Such genetic regions are likely to have contributed to shared adaptations.

This novel approach allowed the researchers to tackle long-standing questions surrounding adaptation from a genomic perspective. First, they identified genomic regions that share similar patterns of sequence variation (characteristic of parallel evolution), as opposed to different patterns of sequence variation (characteristic of convergent evolution), in ecologically similar but geographically separate populations. They conclude that the same gene regions are “often” involved in parallel adaptation — at a conservative estimate, approximately 150 genomic regions are shared among freshwater populations, covering 0.2% of the genome. However, the exact proportion of parallel changes relative to convergent changes cannot be estimated, because the methods the authors used to identify genomic regions of interest rely on their repeated occurrence across populations, such that some regions contributing to convergent adaptation (via different or novel mutations) go unnoticed. So this specific question remains open.

A second, related, question in the study of rapid adaptation is the relative reliance on pre-existing genetic variation compared with new mutations<sup>6</sup>. To address this, Jones *et al.*<sup>4</sup> focused on a single stream in Scotland where marine and freshwater sticklebacks meet and interbreed. They estimate that around 35% of the genomic regions that differ between this population pair (sampled from opposite ends of the hybrid zone in the river) overlap with regions that show an ancient shared origin in the worldwide population samples. The remaining regions of high genetic divergence may be attributable to local adaptations specific to this location or to evolution from different novel mutations. Thus, it seems that repeated evolution of traits may often, but not always, arise from genetic variation that already existed in an ancestral population.

The third question tackled by the authors may be considered a contentious one: that of the relative role of mutations in regulatory versus coding sequences in adaptive change. Whereas some researchers have argued for the predominance of regulatory mutations<sup>7,8</sup>,



**Figure 1 | Adapting to new surroundings.** Jones *et al.*<sup>4</sup> provide a high-quality whole-genome sequence of a freshwater threespine stickleback from Alaska (location indicated by the blue star). This sequence is at 9× coverage, which means that each DNA nucleotide is represented, on average, nine times in the sequence reads used to construct the genome sequence. The authors also generated lower-coverage (2.3×) sequences of stickleback fish from 20 additional populations around the globe — 10 in freshwater (blue circles) and 10 in marine environments (red circles). Their sequence comparisons allow an analysis of how similar genetic adaptations can arise repeatedly in isolated populations.



- 592–603 (2008).  
 12. Rini, B. I. & Atkins, M. B. *Lancet Oncol.* **10**, 992–1000 (2009).  
 13. Páez-Ribes, M. *et al. Cancer Cell* **15**, 220–231 (2009).  
 14. Ebos, J. M. L. *et al. Cancer Cell* **15**, 232–239 (2009).

15. Charafe-Jauffret, E. *et al. Clin. Cancer Res.* **16**, 45–55 (2010).  
 16. Folkins, C. *et al. Cancer Res.* **67**, 3560–3564 (2007).  
 17. Korkaya, H. *et al. PLoS Biol.* **7**, e1000121 (2009).  
 18. Casazza, A. *et al. EMBO Mol. Med.* **4**, 234–250 (2012).

## GENOMICS

# Stickleback is the catch of the day

Whole-genome sequences from a marine fish that has adjusted to life in fresh water give hints about general genetic mechanisms that drive the evolution of adaptations to new environmental niches. [SEE ARTICLE P.55](#)

HOPI E. HOEKSTRA

The traditional recipe for publication of a genome sequence goes something like this: one part ‘biology’ (an often flowery description of the distinctive aspects of the organism whose genome has been sequenced); two parts ‘assembly and annotation’ (how the latest DNA sequencing and computational technologies were used to produce a high-quality sequence); and three parts ‘comparative analyses’ (for example, observations of rapidly evolving genes, or expansion or loss of gene families). All this is followed by a dash of ‘fun speculation’ on how these distinctive genomic characteristics might yield insight into the biology of the creature under study.

Although each genome paper has its own twist, such as filling in a critical gap in the tree of life<sup>1</sup>, a novel sequencing and assembly strategy<sup>2</sup>, or proof that an entire genome can be sequenced by a single laboratory<sup>3</sup>, the

main contribution of such work is usually the sequence itself — as a tool to delve deeper into biological questions. But there are new cooks in the kitchen, and some of them are following a different recipe. On page 55 of this issue, Jones *et al.*<sup>4</sup> report the first whole-genome sequence of the threespine stickleback fish, and at the same time reel in answers to some big questions in evolutionary biology.

The ancestral form of the threespine stickleback (*Gasterosteus aculeatus*) was a marine fish with protective bony ‘armour’ that repeatedly colonized lakes and streams around the world following the retreat of glaciers at the end of the last ice age some 10,000 years ago. In their new freshwater habitats, these fish repeatedly evolved new morphological and physiological adaptations, such as loss of the bony plating and spines and a change in salinity tolerance, and this evolutionary pattern has elevated the threespine stickleback to ‘supermodel’ status<sup>5</sup> in studies of adaptation and speciation. The

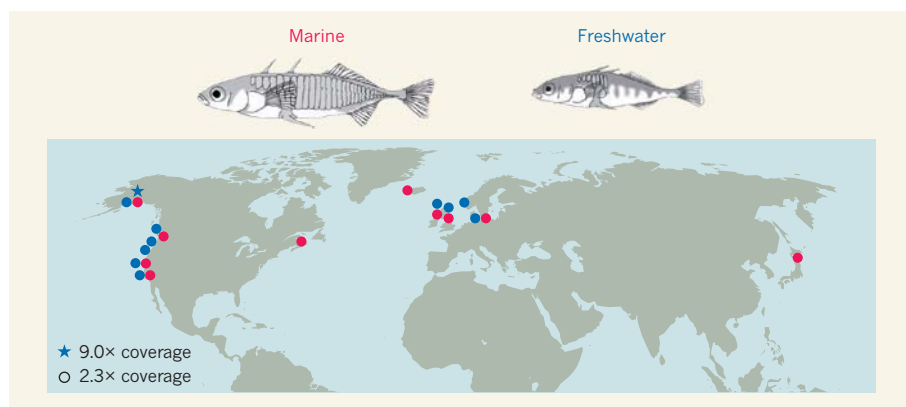
one missing piece has been the complete genome sequence. Jones and colleagues<sup>4</sup> provide just this, with a high-quality sequence of a freshwater Alaskan stickleback.

But the authors don’t stop there. They also report sequence information from another 20 sticklebacks sampled around the globe (Fig. 1) that form 10 geographically linked pairs of marine and freshwater fish. And they develop two complementary approaches — one based on phylogenetic relationships and a second on pairwise genetic distance — to scan the genomes for regions that are substantially different between marine and freshwater sticklebacks but similar within marine, or freshwater, populations from different geographical locations. Such genetic regions are likely to have contributed to shared adaptations.

This novel approach allowed the researchers to tackle long-standing questions surrounding adaptation from a genomic perspective. First, they identified genomic regions that share similar patterns of sequence variation (characteristic of parallel evolution), as opposed to different patterns of sequence variation (characteristic of convergent evolution), in ecologically similar but geographically separate populations. They conclude that the same gene regions are “often” involved in parallel adaptation — at a conservative estimate, approximately 150 genomic regions are shared among freshwater populations, covering 0.2% of the genome. However, the exact proportion of parallel changes relative to convergent changes cannot be estimated, because the methods the authors used to identify genomic regions of interest rely on their repeated occurrence across populations, such that some regions contributing to convergent adaptation (via different or novel mutations) go unnoticed. So this specific question remains open.

A second, related, question in the study of rapid adaptation is the relative reliance on pre-existing genetic variation compared with new mutations<sup>6</sup>. To address this, Jones *et al.*<sup>4</sup> focused on a single stream in Scotland where marine and freshwater sticklebacks meet and interbreed. They estimate that around 35% of the genomic regions that differ between this population pair (sampled from opposite ends of the hybrid zone in the river) overlap with regions that show an ancient shared origin in the worldwide population samples. The remaining regions of high genetic divergence may be attributable to local adaptations specific to this location or to evolution from different novel mutations. Thus, it seems that repeated evolution of traits may often, but not always, arise from genetic variation that already existed in an ancestral population.

The third question tackled by the authors may be considered a contentious one: that of the relative role of mutations in regulatory versus coding sequences in adaptive change. Whereas some researchers have argued for the predominance of regulatory mutations<sup>7,8</sup>,



**Figure 1 | Adapting to new surroundings.** Jones *et al.*<sup>4</sup> provide a high-quality whole-genome sequence of a freshwater threespine stickleback from Alaska (location indicated by the blue star). This sequence is at 9× coverage, which means that each DNA nucleotide is represented, on average, nine times in the sequence reads used to construct the genome sequence. The authors also generated lower-coverage (2.3×) sequences of stickleback fish from 20 additional populations around the globe — 10 in freshwater (blue circles) and 10 in marine environments (red circles). Their sequence comparisons allow an analysis of how similar genetic adaptations can arise repeatedly in isolated populations.

others caution that the role of coding changes may not be negligible<sup>9</sup>. Although these conclusions are not mutually exclusive, data to support either of these claims have previously come predominantly from case studies (of variable quality), rather than from a systematic empirical survey (although see ref. 10 for a review of the latter approach).

Now, however, because their novel mapping approach results in the identification of narrow genomic regions (of the order of only 5 kilobases), Jones and colleagues<sup>4</sup> were able to simply assess whether or not a region implicated in an adaptation contains a coding sequence. If not, then the causal mutation was considered 'regulatory', as they found for 41% of cases. If the region contains both coding and non-coding regions but there is no consistent amino-acid difference between marine and freshwater populations, this was considered 'likely to be regulatory' (43% of cases), but if they do have consistent amino-acid differences the authors classed them as 'coding' (17% of cases). Although some researchers may not

be convinced of these allocations until precise mutations are identified and functionally verified, this large-scale analysis suggests that both amino-acid and regulatory mutations contribute to adaptation, and that most are regulatory in nature.

This genome-wide view of the evolutionary processes acting in sticklebacks has provided us with many clues — that rapid adaptation is often caused by parallel genetic changes, often evolves from pre-existing genetic variation, and often involves regulatory mutations.

Although these results do not provide simple yes or no answers to such concepts, that was not to be expected, as evolution doesn't follow strict rules. Our focus must now shift to 'why' questions. Why are the same genes sometimes used repeatedly in adaptation but at other times involve different genes? Why, in some cases, does evolution take advantage of pre-existing variation and other times new mutations? Why are regulatory mutations sometimes preferred over amino-acid changes and vice versa? Such enquiry will further

improve our understanding of the evolutionary process, and increase our ability to predict evolutionary outcomes. So, for evolutionary geneticists, there are still big fish to fry. ■

**Hopi E. Hoekstra** is in the Department of Organismic and Evolutionary Biology and the Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts 02138, USA.  
e-mail: hoekstra@oeb.harvard.edu

1. Warren, W. C. *et al.* *Nature* **453**, 175–183 (2008).
2. Li, R. *et al.* *Nature* **463**, 311–317 (2010).
3. Zhan, S., Merlin, C., Boore, J. L. & Reppert, S. M. *Cell* **147**, 1171–1185 (2011).
4. Jones, F. C. *et al.* *Nature* **484**, 55–61 (2012).
5. Gibson, G. *Science* **307**, 1890–1891 (2005).
6. Barrett, R. D. H. & Schluter, D. *Trends Ecol. Evol.* **23**, 38–44 (2008).
7. Carroll, S. B. *Cell* **134**, 25–36 (2008).
8. Stern, D. L. & Orgogozo, V. *Evolution* **62**, 2155–2177 (2008).
9. Hoekstra, H. E. & Coyne, J. A. *Evolution* **61**, 995–1016 (2007).
10. Frazier, H. B. *BioEssays* **33**, 469–477 (2011).

## QUANTUM OPTICS

# An entangled walk of photons

**By harnessing the quantum nature of light and guiding the light through a network of circuits integrated in a glass chip, it is possible to mimic fundamental particles undergoing a quantum walk.**

JONATHAN C. F. MATTHEWS  
& MARK G. THOMPSON

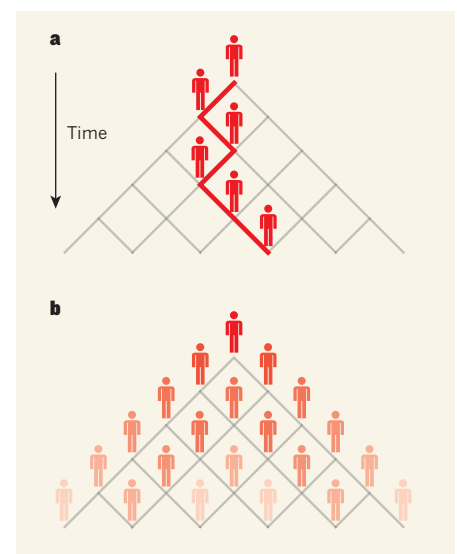
Quantum mechanics is the most successful model of nature that we have, accurately describing fundamental physical processes. Although quantum effects such as entanglement and superposition are counterintuitive and often described as 'spooky', they are being observed and characterized in laboratories worldwide. Writing in *Physical Review Letters*, Sansoni *et al.*<sup>1</sup> describe how they have designed and built an optical network, integrated in a glass chip, that manipulates photons to simulate a process known as a discrete-time quantum walk. Furthermore, using a particular kind of entanglement, the authors simulated different classes of fundamental particles undergoing the quantum-walk process<sup>2</sup>. The results are a step towards the development of quantum-mechanical machines that promise to outperform conventional supercomputers, which operate according to the laws of classical physics\*.

\*This News & Views article was published online on 21 March 2012.

Perhaps the most far-reaching applications of quantum technologies will be based on the American physicist Richard Feynman's proposal that an efficient way to simulate one quantum system is to use another. Such simulators could be used to study complex quantum systems that are computationally hard to simulate or difficult to physically control. We are currently witnessing initial demonstrations of these simulators in the form of purpose-built quantum devices that mimic other, less readily accessible quantum systems.

Developments in quantum simulation are still at the stage of mimicking systems simple enough to be handled with classical computers. However, photonics is an attractive candidate for reaching the point at which quantum machines could outperform state-of-the-art supercomputers for particular tasks. This is partly thanks to the complexity and stability of quantum networks realized with integrated optics, and to the nature of multiple identical photons interfering in a sufficiently complicated optical network<sup>3</sup>.

In classical physics, the analogue of Sansoni and colleagues' discrete-time quantum walk



**Figure 1 | Classical and quantum walks.** **a**, In a classical random walk, a walker must make a choice (randomly) of moving either left or right at each step. After many trials of a fixed number of steps, the walker is most probably found close to its initial starting point. **b**, In a quantum walk, the walker uses a 'quantum coin' mechanism that allows it to move in a superposition of both left and right. The probability distribution of its position after many steps is therefore starkly different from that of a classical random walk, with the walker most likely to be found far from its initial starting point.

would be a form of random walk, a tool already used in a broad range of fields from animal behaviour to economics, for example in modelling the path of a foraging animal or fluctuating prices in the stock market. The simplest random walk is on a line, and is conducted by repeatedly flipping a coin and walking left if



others caution that the role of coding changes may not be negligible<sup>9</sup>. Although these conclusions are not mutually exclusive, data to support either of these claims have previously come predominantly from case studies (of variable quality), rather than from a systematic empirical survey (although see ref. 10 for a review of the latter approach).

Now, however, because their novel mapping approach results in the identification of narrow genomic regions (of the order of only 5 kilobases), Jones and colleagues<sup>4</sup> were able to simply assess whether or not a region implicated in an adaptation contains a coding sequence. If not, then the causal mutation was considered 'regulatory', as they found for 41% of cases. If the region contains both coding and non-coding regions but there is no consistent amino-acid difference between marine and freshwater populations, this was considered 'likely to be regulatory' (43% of cases), but if they do have consistent amino-acid differences the authors classed them as 'coding' (17% of cases). Although some researchers may not

be convinced of these allocations until precise mutations are identified and functionally verified, this large-scale analysis suggests that both amino-acid and regulatory mutations contribute to adaptation, and that most are regulatory in nature.

This genome-wide view of the evolutionary processes acting in sticklebacks has provided us with many clues — that rapid adaptation is often caused by parallel genetic changes, often evolves from pre-existing genetic variation, and often involves regulatory mutations.

Although these results do not provide simple yes or no answers to such concepts, that was not to be expected, as evolution doesn't follow strict rules. Our focus must now shift to 'why' questions. Why are the same genes sometimes used repeatedly in adaptation but at other times involve different genes? Why, in some cases, does evolution take advantage of pre-existing variation and other times new mutations? Why are regulatory mutations sometimes preferred over amino-acid changes and vice versa? Such enquiry will further

improve our understanding of the evolutionary process, and increase our ability to predict evolutionary outcomes. So, for evolutionary geneticists, there are still big fish to fry. ■

**Hopi E. Hoekstra** is in the Department of Organismic and Evolutionary Biology and the Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts 02138, USA.  
e-mail: hoekstra@oeb.harvard.edu

1. Warren, W. C. *et al.* *Nature* **453**, 175–183 (2008).
2. Li, R. *et al.* *Nature* **463**, 311–317 (2010).
3. Zhan, S., Merlin, C., Boore, J. L. & Reppert, S. M. *Cell* **147**, 1171–1185 (2011).
4. Jones, F. C. *et al.* *Nature* **484**, 55–61 (2012).
5. Gibson, G. *Science* **307**, 1890–1891 (2005).
6. Barrett, R. D. H. & Schluter, D. *Trends Ecol. Evol.* **23**, 38–44 (2008).
7. Carroll, S. B. *Cell* **134**, 25–36 (2008).
8. Stern, D. L. & Orgogozo, V. *Evolution* **62**, 2155–2177 (2008).
9. Hoekstra, H. E. & Coyne, J. A. *Evolution* **61**, 995–1016 (2007).
10. Frazier, H. B. *BioEssays* **33**, 469–477 (2011).

## QUANTUM OPTICS

# An entangled walk of photons

By harnessing the quantum nature of light and guiding the light through a network of circuits integrated in a glass chip, it is possible to mimic fundamental particles undergoing a quantum walk.

JONATHAN C. F. MATTHEWS  
& MARK G. THOMPSON

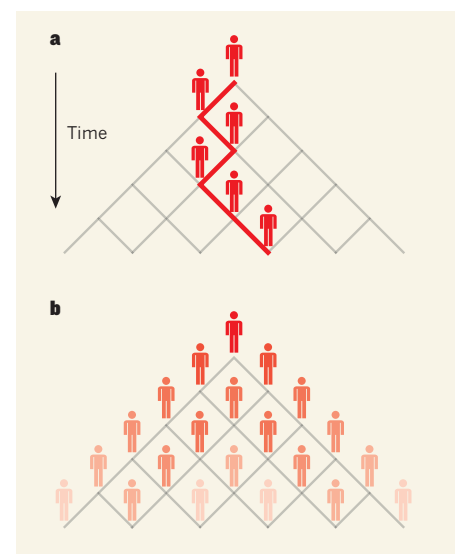
Quantum mechanics is the most successful model of nature that we have, accurately describing fundamental physical processes. Although quantum effects such as entanglement and superposition are counterintuitive and often described as 'spooky', they are being observed and characterized in laboratories worldwide. Writing in *Physical Review Letters*, Sansoni *et al.*<sup>1</sup> describe how they have designed and built an optical network, integrated in a glass chip, that manipulates photons to simulate a process known as a discrete-time quantum walk. Furthermore, using a particular kind of entanglement, the authors simulated different classes of fundamental particles undergoing the quantum-walk process<sup>2</sup>. The results are a step towards the development of quantum-mechanical machines that promise to outperform conventional supercomputers, which operate according to the laws of classical physics\*.

\*This News & Views article was published online on 21 March 2012.

Perhaps the most far-reaching applications of quantum technologies will be based on the American physicist Richard Feynman's proposal that an efficient way to simulate one quantum system is to use another. Such simulators could be used to study complex quantum systems that are computationally hard to simulate or difficult to physically control. We are currently witnessing initial demonstrations of these simulators in the form of purpose-built quantum devices that mimic other, less readily accessible quantum systems.

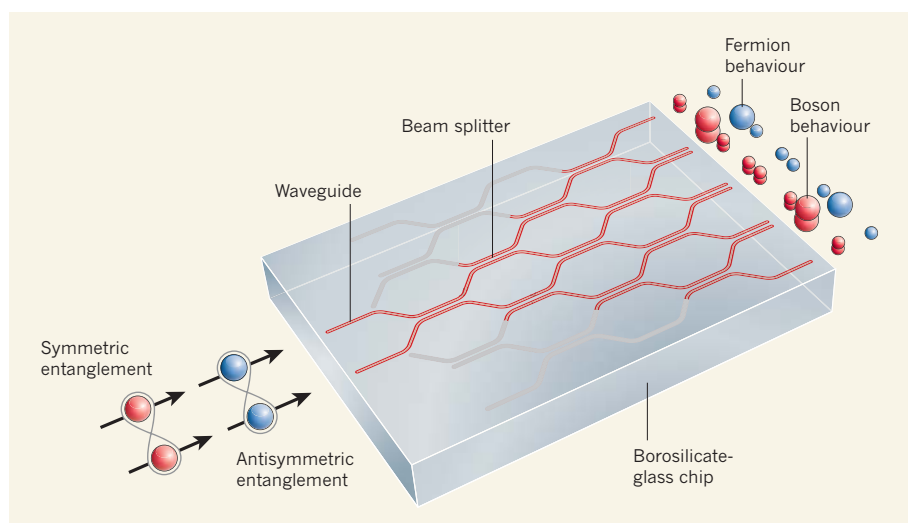
Developments in quantum simulation are still at the stage of mimicking systems simple enough to be handled with classical computers. However, photonics is an attractive candidate for reaching the point at which quantum machines could outperform state-of-the-art supercomputers for particular tasks. This is partly thanks to the complexity and stability of quantum networks realized with integrated optics, and to the nature of multiple identical photons interfering in a sufficiently complicated optical network<sup>3</sup>.

In classical physics, the analogue of Sansoni and colleagues' discrete-time quantum walk



**Figure 1 | Classical and quantum walks.** **a**, In a classical random walk, a walker must make a choice (randomly) of moving either left or right at each step. After many trials of a fixed number of steps, the walker is most probably found close to its initial starting point. **b**, In a quantum walk, the walker uses a 'quantum coin' mechanism that allows it to move in a superposition of both left and right. The probability distribution of its position after many steps is therefore starkly different from that of a classical random walk, with the walker most likely to be found far from its initial starting point.

would be a form of random walk, a tool already used in a broad range of fields from animal behaviour to economics, for example in modelling the path of a foraging animal or fluctuating prices in the stock market. The simplest random walk is on a line, and is conducted by repeatedly flipping a coin and walking left if



**Figure 2 | Discrete quantum walk on a photonic chip.** Sansoni *et al.*<sup>1</sup> fabricated a network of ten waveguide beam splitters in a borosilicate-glass chip using a focused pulsed laser (not shown) and launched two entangled photons (spheres) into the network. The photons underwent a quantum walk and were detected at the chip's output. The output depends on the symmetry of the entanglement between the photons. A symmetric entangled input leads to boson-like behaviour (the photons can occupy the same waveguide and so can exit the waveguides together), whereas an antisymmetrically entangled input yields fermion-like behaviour (the photons cannot occupy the same waveguide and therefore exit through separate waveguides). The size of the spheres at the output illustrates the probability of the photons leaving each of the waveguides, with a bigger size corresponding to a larger probability.

the coin lands on heads, and right if it lands on tails — on average, the walker remains close to where it started. By contrast, if the walker is endowed with quantum properties (for example, in the case of a photon), it moves in a superposition of both left and right at each coin flip, with these multiple possibilities undergoing interference and giving rise to completely different behaviour from that of a classical random walk (Fig. 1).

Such quantum walks were initially conceived as an abstract computer-science model for developing ideas in quantum computation. But now researchers are interested in using quantum walks to mimic real systems that exhibit local coupling. Examples of such systems are materials in which particles with spin are coupled or excitons travelling around large molecules (an exciton is a particle-like entity formed by an electron and the 'hole' left by a missing electron).

So far, all demonstrations of quantum walks (including those based on cold atoms, nuclear magnetic resonance and photonics) have sidestepped the requirement of a quantum computer by constructing purpose-built systems that exhibit quantum-walk dynamics. These demonstrations have allowed tests of quantum phenomena to be performed ahead of building a full-scale quantum computer. The first walks to be realized were of single quantum walkers, which display classical wave-like behaviour<sup>4</sup>. Multiple quantum walkers lead to behaviour that cannot be explained with wave phenomena alone, and have been reported<sup>5</sup> using two indistinguishable photons launched into integrated optical circuits. Sansoni and colleagues'

demonstration<sup>1</sup> goes further by entangling the walkers in a discrete-time quantum-walk network<sup>6</sup> achieved using an integrated network of waveguide beam splitters.

The particles of choice for Sansoni *et al.* are photons, which show great promise for quantum technologies<sup>7</sup>. They are readily manipulated at the single-photon level and experience negligible noise — in effect, they 'leak' very little quantum information to the environment. Working with photons also allows the deployment of integrated quantum photonics<sup>8</sup> to make reconfigurable, complex miniature circuits that are inherently stable within fractions of an optical wavelength (about 10 nanometres)<sup>9</sup>. The authors fabricated their circuits using a laser 'direct-write' technique that allows three-dimensional optical circuits to be written into a solid block of borosilicate glass using a tightly focused beam of laser pulses<sup>10</sup>.

Sansoni and colleagues' optical network<sup>1</sup> consisted of ten waveguide beam splitters at discrete points in a centimetre-scale glass circuit. Each beam splitter provided the 'quantum coin' mechanism that repeatedly placed the photons in quantum superposition as they progressed through the network (Fig. 2). Moreover, thanks to the type of entanglement they used and the fact that their device preserves photon polarization, the authors were able to simulate quantum walks not just for photons (and therefore the boson class of particles) but also fundamentally different classes of particles (the fermion and intermediate-type families). This approach can be generalized for any quantum network (not just quantum walks) and can be scaled-up to simulate any number of particles<sup>2</sup>.

The scalability of integrated-optics circuits allows for incredible flexibility: waveguide networks, consisting of hundreds of waveguides and beam splitters, could be implemented in future experiments to simulate more complex dynamics than those demonstrated by the authors. So far, only two non-interacting quantum particles have been simulated undergoing relatively simple quantum walks. Large numbers of photons will be required for any practical quantum-photonics technology, and future developments could include the simulation of interacting quantum particles (such as charged particles) through implementation of photon-photon nonlinearities<sup>7</sup> using auxiliary photon measurement or light-matter interaction, for example.

However, a recent, and perhaps surprising, theoretical result<sup>3</sup> is that a large number of identical photons themselves pose computationally difficult problems due to their bosonic nature, even in the absence of interaction. Using just tens of photons guided in a circuit with hundreds of waveguides could lead to the first quantum computation competitive with the equivalent calculation performed with the best classical computers currently available. Sansoni and colleagues' study is a development in this direction, and paves the way for using integrated quantum circuits to build a practical quantum simulator. ■

**Jonathan C. F. Matthews and Mark G. Thompson** are at the Centre for Quantum Photonics, H. H. Wills Physics Laboratory and the Department of Electrical and Electronic Engineering, University of Bristol, Bristol BS8 1UB, UK.  
e-mail: mark.thompson@bristol.ac.uk

1. Sansoni, L. *et al.* *Phys. Rev. Lett.* **108**, 010502 (2012).
2. Matthews, J. C. F. *et al.* preprint at [arxiv.org/abs/1106.1166v1](https://arxiv.org/abs/1106.1166v1) (2011).
3. Aaronson, S. & Arkhipov, A. *Proc. 43rd Annu. ACM Symp. Theory Comput.* 333–342 (2011).
4. Knight, P. L., Roldán, E. & Sipe, J. E. *Phys. Rev. A* **68**, 020301(R) (2003).
5. Peruzzo, A. *et al.* *Science* **329**, 1500–1503 (2010).
6. Omar, Y., Paunković, N., Sheridan, L. & Bose, S. *Phys. Rev. A* **74**, 042304 (2006).
7. O'Brien, J. L., Furusawa, A. & Vučković, J. *Nature Photon.* **3**, 687–695 (2009).
8. Politi, A. *et al.* *Science* **320**, 646–649 (2008).
9. Shadbolt, P. *et al.* *Nature Photon.* **6**, 45–49 (2012).
10. Sansoni, L. *et al.* *Phys. Rev. Lett.* **105**, 200503 (2010).

#### CORRECTION

In the News & Views article 'Molecular biology: How to duplicate a DNA package' by Alysia Vandenberg & Geneviève Almouzni (*Nature* **483**, 412–413; 2012), reference 5 should have been cited in the first paragraph, rather than reference 1. And in the sixth paragraph, reference 5 should not have been cited; instead, reference 1 should have been cited after "chromatin remodeller enzymes".



# Global warming preceded by increasing carbon dioxide concentrations during the last deglaciation

Jeremy D. Shakun<sup>1,2</sup>, Peter U. Clark<sup>3</sup>, Feng He<sup>4</sup>, Shaun A. Marcott<sup>3</sup>, Alan C. Mix<sup>3</sup>, Zhengyu Liu<sup>4,5,6</sup>, Bette Otto-Bliesner<sup>7</sup>, Andreas Schmittner<sup>3</sup> & Edouard Bard<sup>8</sup>

The covariation of carbon dioxide (CO<sub>2</sub>) concentration and temperature in Antarctic ice-core records suggests a close link between CO<sub>2</sub> and climate during the Pleistocene ice ages. The role and relative importance of CO<sub>2</sub> in producing these climate changes remains unclear, however, in part because the ice-core deuterium record reflects local rather than global temperature. Here we construct a record of global surface temperature from 80 proxy records and show that temperature is correlated with and generally lags CO<sub>2</sub> during the last (that is, the most recent) deglaciation. Differences between the respective temperature changes of the Northern Hemisphere and Southern Hemisphere parallel variations in the strength of the Atlantic meridional overturning circulation recorded in marine sediments. These observations, together with transient global climate model simulations, support the conclusion that an antiphased hemispheric temperature response to ocean circulation changes superimposed on globally in-phase warming driven by increasing CO<sub>2</sub> concentrations is an explanation for much of the temperature change at the end of the most recent ice age.

Understanding the causes of the Pleistocene ice ages has been a significant question in climate dynamics since they were discovered in the mid-nineteenth century. The identification of orbital frequencies in the marine <sup>18</sup>O/<sup>16</sup>O record, a proxy for global ice volume, in the 1970s demonstrated that glacial cycles are ultimately paced by astronomical forcing<sup>1</sup>. Initial measurements of air bubbles in Antarctic ice cores in the 1980s revealed that greenhouse gas concentrations also increased and decreased over the last glacial cycle<sup>2,3</sup>, suggesting they too may be part of the explanation. The ice-core record now extends back 800,000 yr and shows that local Antarctic temperature was strongly correlated with and seems to have slightly led changes in CO<sub>2</sub> concentration<sup>4</sup>. The implication of this relationship for understanding the role of CO<sub>2</sub> in glacial cycles, however, remains unclear. For instance, proxy data have variously been interpreted to suggest that CO<sub>2</sub> was the primary driver of the ice ages<sup>5</sup>, a more modest feedback on warming<sup>6,7</sup> or, perhaps, largely a consequence rather than cause of past climate change<sup>8</sup>. Similarly, although climate models generally require greenhouse gases to explain globalization of the ice-age signal, they predict a wide range (one-third to two-thirds) in the contribution of greenhouse gases to ice-age cooling, with additional contributions from ice albedo and other effects<sup>9,10</sup>. Moreover, models have generally used prescribed forcings to simulate snapshots in time and thus by design do not distinguish the timing of changes in various forcings relative to responses.

Global temperature reconstructions and transient model simulations spanning the past century and millennium have been essential to the attribution of recent climate change, and a similar strategy would probably improve our understanding of glacial cycle dynamics. Here we use a network of proxy temperature records that provide broad spatial coverage to show that global temperature closely tracked the

increase in CO<sub>2</sub> concentration over the last deglaciation, and that variations in the Atlantic meridional overturning circulation (AMOC) caused a seesawing of heat between the hemispheres, supporting an early hypothesis that identified potentially important roles for these mechanisms<sup>11</sup>. These findings, supported by transient simulations with a coupled ocean–atmosphere general circulation model, can explain the lag of CO<sub>2</sub> behind Antarctic temperature in the ice-core record and are consistent with an important role for CO<sub>2</sub> in driving global climate change over glacial cycles.

## Global temperature

We calculate the area-weighted mean of 80 globally distributed, high-resolution proxy temperature records to reconstruct global surface temperature during the last deglaciation (Methods and Fig. 1). The global temperature stack shows a two-step rise, with most warming occurring during and right after the Oldest Dryas and Younger Dryas intervals and relatively little temperature change during the Last Glacial Maximum (LGM), the Bølling–Allerød interval and the early Holocene epoch (Fig. 2a). The atmospheric CO<sub>2</sub> record from the EPICA Dome C ice core<sup>12</sup>, which has recently been placed on a more accurate timescale<sup>13</sup>, has a similar two-step structure and is strongly correlated with the temperature stack ( $r^2 = 0.94$  (coefficient of determination),  $P = 0.03$ ; Fig. 2a).

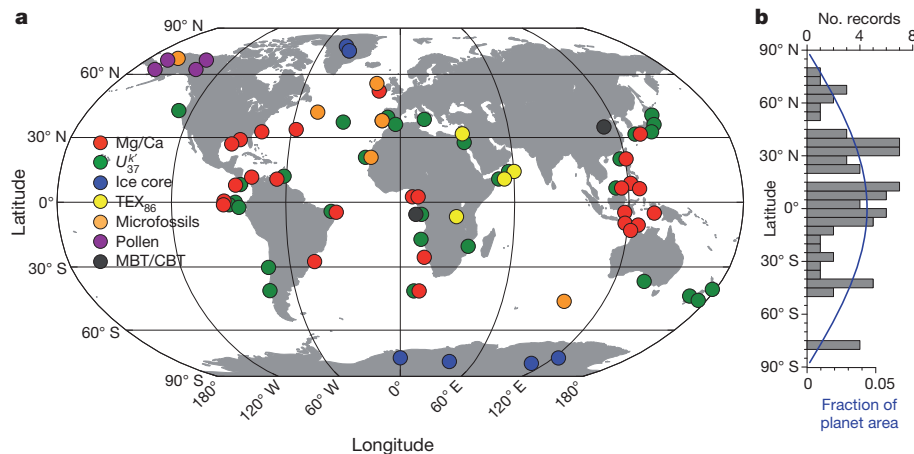
Lag correlations quantify the timing of change in the temperature stack relative to CO<sub>2</sub> from 20–10 kyr ago, an interval that spans the period during which low LGM CO<sub>2</sub> concentrations increased to almost pre-industrial values. Our results indicate that CO<sub>2</sub> probably leads global warming over the course of the deglaciation (Fig. 2b). A comparison of the global temperature stack with Antarctic temperature provides further support for this relative timing, in showing that

<sup>1</sup>Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts 02138, USA. <sup>2</sup>Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York 10964, USA.

<sup>3</sup>College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, Oregon 97331, USA. <sup>4</sup>Center for Climatic Research, University of Wisconsin, Madison, Wisconsin 53706, USA.

<sup>5</sup>Department of Atmospheric and Oceanic Sciences, University of Wisconsin, Madison, Wisconsin 53706, USA. <sup>6</sup>Laboratory for Ocean-Atmosphere Studies, Peking University, Beijing 100871, China.

<sup>7</sup>Climate and Global Dynamics Division, National Center for Atmospheric Research, Boulder, Colorado 80307-3000, USA. <sup>8</sup>CEREGE, Collège de France, CNRS-Université Aix-Marseille, Europole de l'Arbois, 13545 Aix-en-Provence, France.



**Figure 1 | Proxy temperature records.** **a**, Location map. CBT, cyclization ratio of branched tetraethers; MBT, methylation index of branched tetraethers;  $TEX_{86}$ , tetraether index of tetraethers consisting of 86 carbon atoms;

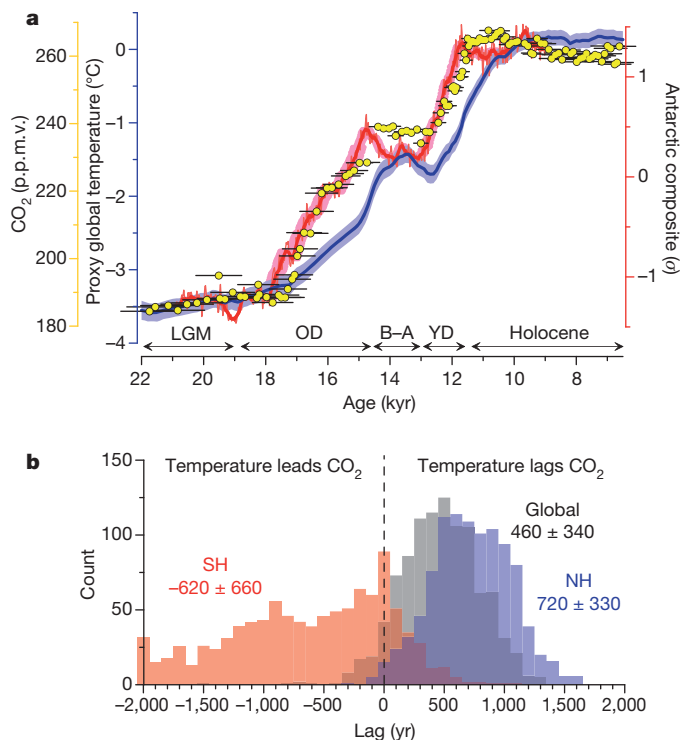
$U_{37}^K$ , alkenone unsaturation index. **b**, Distribution of the records by latitude (grey histogram) and areal fraction of the planet in 5° steps (blue line).

although the structure of the global stack is similar to the pattern of Antarctic temperature change, it lags Antarctica by several centuries to a millennium throughout most of the deglaciation (Fig. 2a). Thus, the small apparent lead of Antarctic temperature over  $CO_2$  in the ice-core records<sup>12,14</sup> does not apply to global temperature. An additional evaluation of this result comes from an objective identification of inflection points in the  $CO_2$  and global temperature records, which

suggests that changes in  $CO_2$  concentration were either synchronous with or led global warming during the various steps of the deglaciation (Supplementary Table 2). An important exception is the onset of deglaciation, which features about 0.3 °C of global warming before the initial increase in  $CO_2 \sim 17.5$  kyr ago. This finding suggests that  $CO_2$  was not the cause of initial warming. We return to this point below. Nevertheless, the overall correlation and phasing of global temperature and  $CO_2$  are consistent with  $CO_2$  being an important driver of global warming during the deglaciation, with the centennial-scale lag of temperature behind  $CO_2$  being consistent with the thermal inertia of the climate system owing to ocean heat uptake and ice melting<sup>15</sup>.

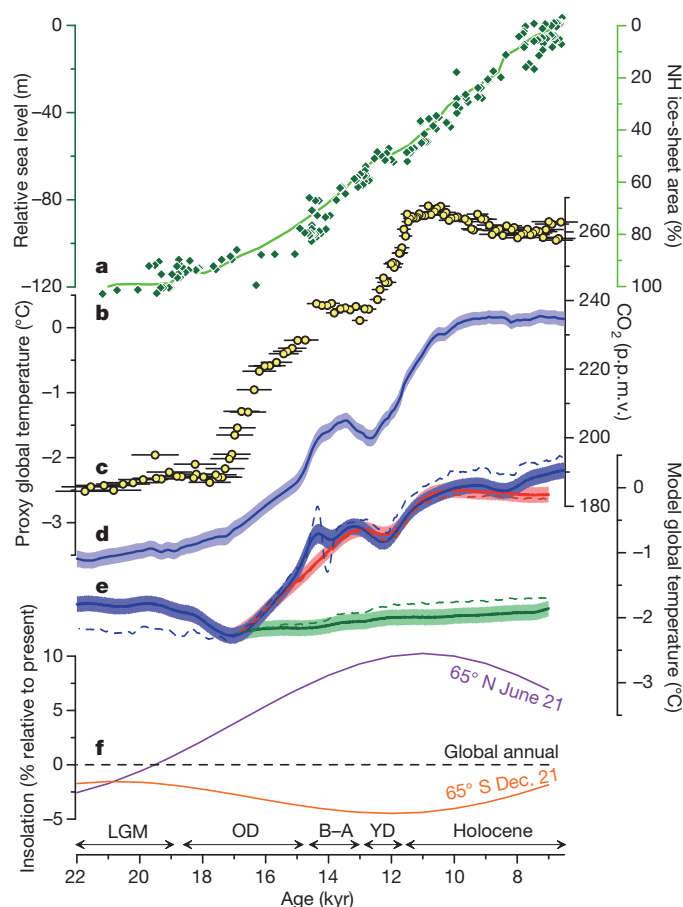
Although other mechanisms contributed to climate change during the ice ages, climate models suggest that their impacts were regional and thus cannot explain the global extent of temperature changes documented by our stacked record alone<sup>9,16,17</sup>. This conclusion is supported by the distinct differences, relative to the temperature stack, in the temporal variabilities of other likely climate change agents (Fig. 3). For example, insolation is a smoothly varying sinusoid that is in antiphase between the hemispheres and sums to near zero globally at the top of the atmosphere (Fig. 3f). Although spatial and temporal asymmetries in albedo could convert insolation to a non-zero forcing at Earth's surface, it is unlikely to account for much of the step-like structure and global nature of the temperature stack.

Similarly, although ice-sheet extent and its associated albedo (from ice cover and emergent continental shelves) and orographic forcing decreased through the deglaciation, global ice volume and area changed only slowly or not at all during intervals of pronounced global warming such as the Oldest Dryas and Younger Dryas, and the greatest volume or area loss in fact occurred during intervals of little or no warming around 19 kyr ago and the Bølling–Allerød (Fig. 3a, b). This distinction is particularly notable during the early Holocene, when the temperature stack had reached interglacial levels while nearly one-third of the excess global ice still remained, although we note that any ice-driven warming would have been partly offset by decreasing greenhouse gas forcing (Fig. 3c and Supplementary Fig. 29a). The apparently small influence of ice-sheet forcing on the temperature stack is consistent with general circulation models that suggest its effect was largely confined to the northern mid to high latitudes and was otherwise modest in the areas sampled by our proxy network<sup>16–18</sup>, which is biased away from the ice sheets. Our results, therefore, do not preclude an important contribution to global mean warming from ice-sheet retreat, but suggest that much of this warming was spatially restricted and may be inherently under-represented owing to the lack of suitable palaeotemperature records from and proximal to areas formerly covered by ice.



**Figure 2 |  $CO_2$  concentration and temperature.** **a**, The global proxy temperature stack (blue) as deviations from the early Holocene (11.5–6.5 kyr ago) mean, an Antarctic ice-core composite temperature record<sup>42</sup> (red), and atmospheric  $CO_2$  concentration (refs 12, 13; yellow dots). The Holocene, Younger Dryas (YD), Bølling–Allerød (B–A), Oldest Dryas (OD) and Last Glacial Maximum (LGM) intervals are indicated. Error bars,  $1\sigma$  (Methods); p.p.m.v., parts per million by volume. **b**, The phasing of  $CO_2$  concentration and temperature for the global (grey), Northern Hemisphere (NH; blue) and Southern Hemisphere (SH; red) proxy stacks based on lag correlations from 20–10 kyr ago in 1,000 Monte Carlo simulations (Methods). The mean and  $1\sigma$  of the histograms are given.  $CO_2$  concentration leads the global temperature stack in 90% of the simulations and lags it in 6%.





**Figure 3 | Global temperature and climate forcings.** **a**, Relative sea level<sup>26</sup> (diamonds). **b**, Northern Hemisphere ice-sheet area (line) derived from summing the extents of the Laurentide<sup>43</sup>, Cordilleran<sup>43</sup> and Scandinavian (R. Gyllencreutz and J. Mangerud, personal communication) ice sheets through time. **c**, Atmospheric CO<sub>2</sub> concentration. **d**, Global proxy temperature stack. **e**, Modelled global temperature stacks from the ALL (blue), CO<sub>2</sub> (red) and ORB (green) simulations. Dashed lines show global mean temperatures in the simulations, using sea surface temperatures over ocean and surface air temperatures over land. **f**, Insolation forcing for latitudes 65° N (purple) and 65° S (orange) at the local summer solstice, and global mean annual insolation (dashed black)<sup>44</sup>. Error bars, 1 $\sigma$ .

Unlike these regional-scale forcings, methane, nitrous oxide and possibly dust are global in nature. Because greenhouse gas forcing was dominated by CO<sub>2</sub> (ref. 19; Supplementary Fig. 29a), and because at the onsets of the Bølling–Allerød, Younger Dryas and Holocene the methane and nitrous oxide records have small step changes like those of the global temperature stack, including these greenhouse gases leaves the correlation with the stack essentially unchanged ( $r^2 = 0.93$ ) and slightly decreases the temperature lag ( $250 \pm 340$  yr) (Supplementary Fig. 29). Global dust forcing is poorly constrained<sup>19</sup>, however, and we cannot dismiss it as a potentially important driver of global temperature independent of greenhouse warming. Vegetation forcing is likewise difficult to assess<sup>19</sup> and may have significantly contributed to global warming. These uncertainties notwithstanding, we suggest that the increase in CO<sub>2</sub> concentration before that of global temperature is consistent with CO<sub>2</sub> acting as a primary driver of global warming, although its continuing increase is presumably a feedback from changes in other aspects of the climate system.

The global temperature lag behind CO<sub>2</sub> identified here relies critically on the chronological accuracy of these records. The largest uncertainty in the proxy age models is associated with radiocarbon reservoir corrections, which affect marine, but generally not terrestrial, records. A recent synthesis found similar temperature variabilities in

land and ocean proxy records during the last deglaciation, but that the timing may be slightly earlier in the marine records<sup>20</sup> (Supplementary Fig. 4). Likewise, the pattern of temperature changes at upwelling sites, where reservoir ages may be more variable, is similar to that at non-upwelling sites but again seems somewhat older (Supplementary Fig. 4). These relationships imply that marine reservoir corrections may have been underestimated, which would shift the temperature stack to later times in some intervals and increase its average lag relative to CO<sub>2</sub>. We also evaluated the EPICA Dome C CO<sub>2</sub> chronology<sup>13</sup> by comparing the Dome C methane record on this timescale with the more precisely dated Greenland composite methane record on the GICC05 timescale<sup>21</sup>. This comparison suggests that the EPICA Dome C CO<sub>2</sub> age model may be one to two centuries too young during parts of the deglaciation (Supplementary Fig. 7), which would further increase the lag of CO<sub>2</sub> over global temperature. We thus regard the lag of global temperature behind CO<sub>2</sub> reported here as conservative.

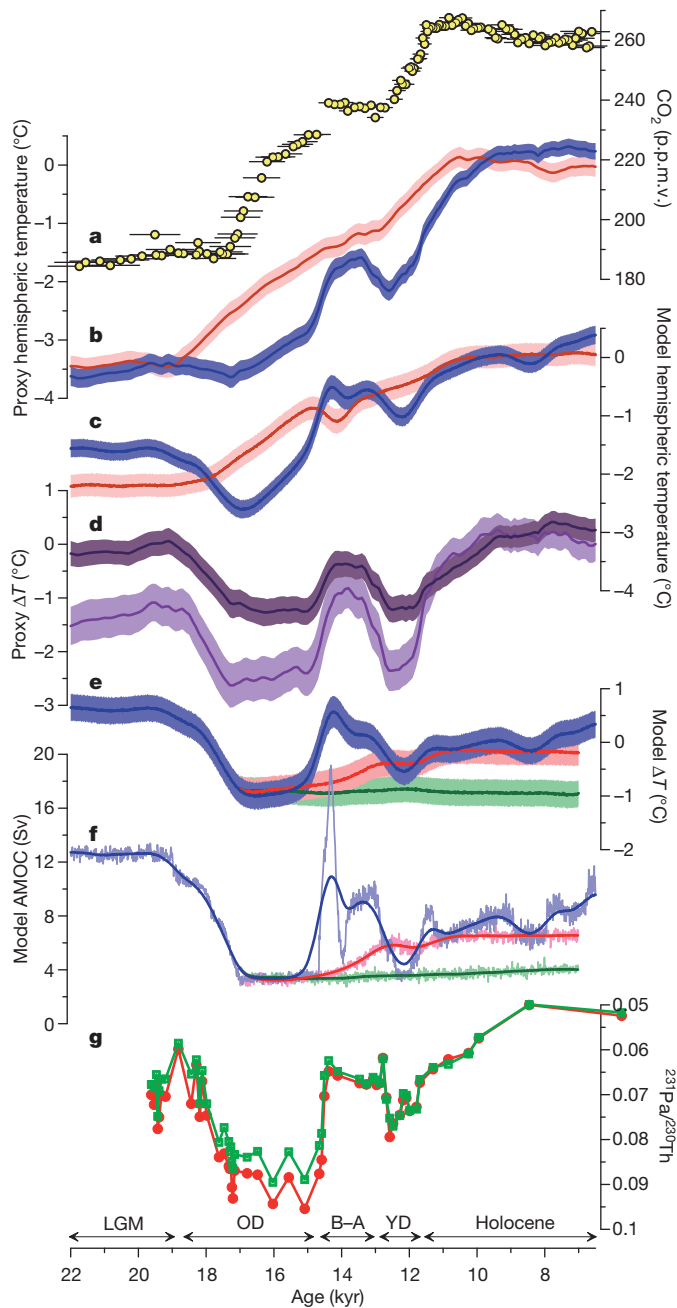
### Hemispheric temperatures

The lead of Antarctic temperature over global temperature indicates spatial variability in the pattern of deglacial warming. To examine this spatial variability further, we calculated separate temperature stacks for the Northern Hemisphere and Southern Hemisphere and found that the magnitude of deglacial warming in the two hemispheric stacks is nearly identical (Fig. 4b). Given that greater LGM cooling probably occurred in the areas affected by the Northern Hemisphere ice sheets<sup>9,17</sup>, this result provides additional support for our inference that the proxy network under-represents the regional impact of the ice sheets. Each hemispheric stack also shows a two-step warming as seen in the global stack and the CO<sub>2</sub> record (Fig. 4a). Otherwise, the hemispheric stacks differ in two main ways. First, lag correlations suggest that whereas Southern Hemisphere temperature probably leads CO<sub>2</sub>, consistent with the Antarctic ice-core results<sup>12</sup>, Northern Hemisphere temperature lags CO<sub>2</sub> (Fig. 2b). Second, the Northern Hemisphere shows modest coolings coincident with the onset of Southern Hemisphere warmings, and the warming steps are concave-up in the north but are concave-down in the south (Fig. 4b).

Calculating the temperature difference,  $\Delta T$ , between the two hemispheric stacks yields an estimate of the heat distribution between the hemispheres, and reveals two large millennial-scale oscillations that are one-quarter to one-third of the glacial–interglacial range in global temperature (Fig. 4d). We attribute the variability in  $\Delta T$  to variations in the strength of the AMOC and its attendant effects on cross-equatorial heat transport<sup>22,23</sup>. A strong correlation of  $\Delta T$  with a kinematic proxy (Pa/Th, the protactinium/thorium ratio) for the strength of the AMOC<sup>24</sup> ( $r^2 = 0.79$ ,  $P = 0.03$ ) supports this interpretation (Fig. 4g). We find that  $\Delta T$  decreases during the Oldest Dryas and Younger Dryas intervals, when the Pa/Th record suggests that the AMOC is weak and heat transfer between the hemispheres is reduced, and that  $\Delta T$  increases during the LGM, the Bølling–Allerød and the Holocene, when the AMOC is stronger and transports heat from the south to the north. Recalculating  $\Delta T$  for Atlantic-only records yields the same relations, but they are more pronounced and better correlated with Pa/Th ( $r^2 = 0.86$ ,  $P = 0.01$ ), as would be expected given the importance of the AMOC in this ocean (Fig. 4d). We note that this seesawing of heat between the hemispheres explains the contrast between the lead of Antarctic temperature over CO<sub>2</sub> and the lag of global (and Northern Hemisphere) temperature behind CO<sub>2</sub>.

### Transient modelling

We evaluate potential physical explanations for the correlations between temperature, CO<sub>2</sub> concentration and AMOC variability in three transient simulations of the last deglaciation using the Community Climate System Model version 3 (CCSM3; ref. 25) of the US National Center for Atmospheric Research. The first simulation (ALL) runs from 22 to 6.5 kyr ago and is driven by changes in greenhouse gases, insolation, ice sheets and freshwater fluxes (the last of which is adjusted iteratively



**Figure 4 | Hemispheric temperatures.** **a**, Atmospheric CO<sub>2</sub> concentration. **b**, Northern Hemisphere (blue) and Southern Hemisphere (red) proxy temperature stacks. **c**, Modelled Northern Hemisphere (blue) and Southern Hemisphere (red) temperature stacks from the ALL simulation. **d**, Northern Hemisphere minus Southern Hemisphere proxy temperature stacks (dark purple). North Atlantic minus South Atlantic region proxy temperature stacks (light purple). **e**, Modelled Northern Hemisphere minus Southern Hemisphere temperature stacks in the ALL (blue), CO<sub>2</sub> (red) and ORB (green) simulations. **f**, Modelled AMOC strength in the ALL (blue), CO<sub>2</sub> (red) and ORB (green) simulations. **g**, North Atlantic sediment core OCE326-GGC5 <sup>231</sup>Pu/<sup>230</sup>Th (ref. 24). Temperatures are given as deviations from the early Holocene (11.5–6.5 kyr ago) mean. Error bars, 1 $\sigma$ .

and thus is a tunable parameter). The second simulation (CO<sub>2</sub>) is forced only by imposed changes in greenhouse gases (CO<sub>2</sub>, methane and nitrous oxide), and the third simulation (ORB) is forced only by orbitally driven insolation variations. All other forcing factors for the second and third simulations, which run from 17 to 7 kyr ago, are held constant at their values at 17 kyr ago. All three simulations include dynamic vegetation feedback and a fixed annual cycle of aerosol

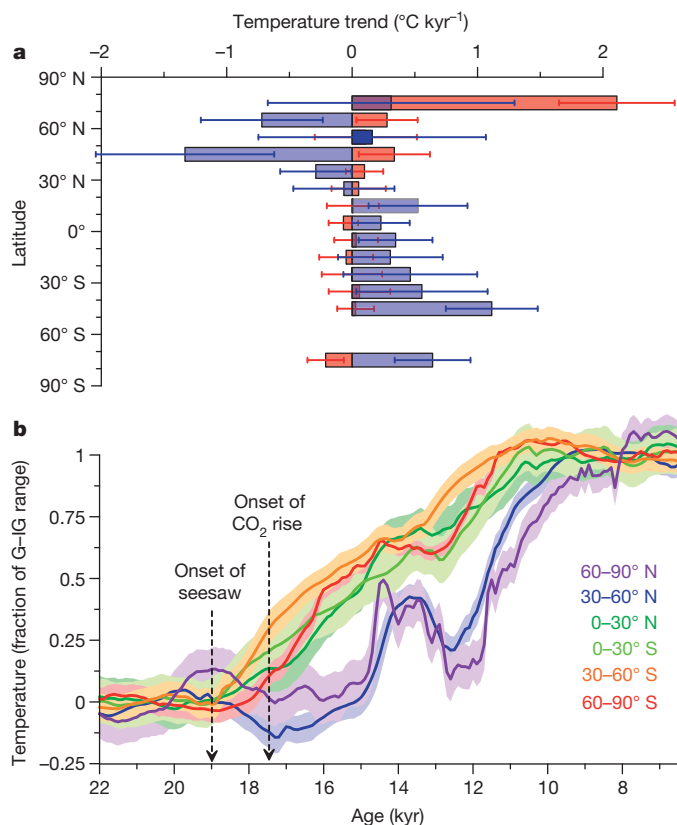
forcing. We sample the model output at the locations of the 80 proxy records, recording sea surface temperatures for the marine records and surface air temperatures for the land records, and stacking these sampled model time series just as we did the data. We also correct sea surface temperature time series from the ALL simulation for sea-level changes by scaling the eustatic sea-level curve<sup>26</sup> to a warming by 0.32 °C at the LGM lowstand<sup>27</sup>. To simulate uncertainties in the model results comparable to those of the data, we generated 1,000 Monte Carlo simulations in which the modelled time series were perturbed with random age-model ( $\pm 300$  yr, 1 $\sigma$ ) and temperature ( $\pm 1$  °C, 1 $\sigma$ ) errors at each site.

The ALL model temperature stack from the 80 sites is similar to the global mean temperature from the model ( $r^2 = 0.97$ ), suggesting that the proxy sites represent the globe fairly well, although the amplitude of warming is slightly smaller in the stack (Fig. 3e and Supplementary Fig. 12). The ALL model stack is also similar to the CO<sub>2</sub> model stack in shape and amplitude ( $r^2 = 0.98$ ; Fig. 3e). Because the CO<sub>2</sub> model stack reflects a response to only greenhouse gas forcing, its similarity to the ALL stack suggests that greenhouse gases can explain most of the mean warming at these 80 sites. The ORB model stack, by contrast, shows only minor warming, consistent with a modest role for orbital forcing in directly driving global temperature changes.

Calculating the difference in model temperature between the Northern Hemisphere and the Southern Hemisphere at the proxy sites in the three simulations yields  $\Delta T$  time series that are strongly correlated with variations in modelled AMOC strength in each simulation ( $r^2 = 0.95$  for ALL, 0.98 for CO<sub>2</sub>; Fig. 4f). Only  $\Delta T$  for the ALL simulation, however, shows millennial-scale variability similar to that seen in the proxy  $\Delta T$  time series and the Pa/Th record (Fig. 4d, e, g). These results suggest that ocean circulation changes driven primarily by freshwater flux, rather than by direct forcing from greenhouse gases or orbits, are plausible causes of the hemispheric differences in temperature change seen in the proxy records. Furthermore, in the ALL simulation the Southern Hemisphere temperature stack leads Northern Hemisphere (and global) temperature during the two deglacial warming steps (Fig. 4c), supporting our inference that AMOC-driven internal heat redistributions explain the Antarctic temperature lead and global temperature lag relative to CO<sub>2</sub>. Lag correlations from 20–10 kyr ago suggest that the modelled global temperature lags CO<sub>2</sub> concentration by 120 yr, which is within the uncertainty range of the proxy-based lag.

### The trigger for deglacial warming

The proxy database provides an opportunity to explore what triggers deglacial warming. Substantial temperature change at all latitudes (Fig. 5b), as well as a net global warming of about 0.3 °C (Fig. 2a), precedes the initial increase in CO<sub>2</sub> concentration at 17.5 kyr ago, suggesting that CO<sub>2</sub> did not initiate deglacial warming. This early global warming occurs in two phases: a gradual increase between 21.5 and 19 kyr ago followed by a somewhat steeper increase between 19 and 17.5 kyr ago (Fig. 2a). The first increase is associated with mean warming of the northern mid to high latitudes, most prominently in Greenland, as there is little change occurring elsewhere at this time (Fig. 5 and Supplementary Fig. 20). The second increase occurs during a pronounced interhemispheric seesaw event (Fig. 5), presumably related to a reduction in AMOC strength, as seen in the Pa/Th record and our modelling (Fig. 4f, g). Tropical and Southern Hemisphere warming seem to have more than offset northern extratropical cooling, however, perhaps as a result of an asymmetry in the response of feedbacks such as Southern Ocean sea ice or tropical water vapour, leading to the global mean response. Alternatively, this non-zero-sum response may reflect proxy biases, as tropical warming is not equally evident in all proxies (Supplementary Fig. 20). In any event, we suggest that these spatiotemporal patterns of temperature change are consistent with warming at northern mid to high latitudes, leading to a reduction in the AMOC at ~19 kyr ago, being the trigger for the



**Figure 5 | Temperature change before increase in CO<sub>2</sub> concentration.**

**a**, Linear temperature trends in the proxy records from 21.5–19 kyr ago (red) and 19–17.5 kyr ago (blue) averaged in 10° latitude bins with 1 $\sigma$  uncertainties. **b**, Proxy temperature stacks for 30° latitude bands with 1 $\sigma$  uncertainties. The stacks have been normalized by the glacial–interglacial (G–IG) range in each time series to facilitate comparison.

global deglacial warming that followed, although more records will be required to confirm the extent and magnitude of early warming at such latitudes. A possible forcing model to explain this sequence of events starts with rising boreal summer insolation driving northern warming<sup>28</sup>. This leads to the observed retreat of Northern Hemisphere ice sheets<sup>26</sup> and the increase in sea level<sup>29</sup> commencing ~19 kyr ago (Fig. 3a, b), with the attendant freshwater forcing causing a reduction in the AMOC that warms the Southern Hemisphere through the bipolar seesaw<sup>30</sup>.

Recent studies of the deglaciation<sup>31,32</sup> have shown a strong correlation between times of minima in the AMOC and maxima in CO<sub>2</sub> release, consistent with our  $\Delta T$  proxy for AMOC strength (Fig. 4d), suggesting that a change in the AMOC may have also contributed to CO<sub>2</sub> degassing from the deep Southern Ocean though its influence on the extent of Southern Ocean sea ice<sup>33</sup>, the position of the southern westerlies<sup>34</sup> or the efficiency of the biological pump<sup>35</sup>. Further insight into this relationship is provided by meridional differences in the timing of proxy temperature change following the reduction in AMOC after ~19 kyr ago. A near-synchronous seesaw response is seen from the high northern latitudes to the mid southern latitudes, whereas strong Antarctic warming and the increase in CO<sub>2</sub> concentration lag the AMOC change<sup>36</sup> (Figs 2a and 5b). This lag suggests that the high-southern-latitude temperature response to an AMOC perturbation may involve a time constant such as that from Southern Ocean thermal inertia<sup>23,37</sup>, whereas the CO<sub>2</sub> response requires a threshold in AMOC reduction to displace southern winds or sea ice sufficiently<sup>38</sup> or to perturb the ocean's biological pump<sup>35</sup>. We also suggest that the delay of Antarctic warming that follows the AMOC seesaw event 19 kyr ago and occurs relative to the mid southern

latitudes over the entire deglaciation (Fig. 5b) is difficult to reconcile with hypotheses invoking a southern high-latitude trigger for deglaciation<sup>39,40</sup>.

Our global temperature stack and transient modelling point to CO<sub>2</sub> as a key mechanism of global warming during the last deglaciation. Furthermore, our results support an interhemispheric seesawing of heat related to AMOC variability and suggest that these internal heat redistributions explain the lead of Antarctic temperature over CO<sub>2</sub> while global temperature was in phase with or slightly lagged CO<sub>2</sub>. Lastly, the global proxy database suggests that parts of the northern mid to high latitudes were the first to warm after the LGM, which could have initiated the reduction in the AMOC that may have ultimately caused the increase in CO<sub>2</sub> concentration.

## METHODS SUMMARY

The data set compiled in this study contains most published high-resolution (median resolution, 200 yr), well-dated ( $n = 636$  radiocarbon dates) temperature records from the last deglaciation (see Supplementary Information for the full database). Sixty-seven records are from the ocean and are interpreted to reflect sea surface temperatures, and the remaining 13 record air or lake temperatures on land. All records span 18–11 kyr ago and ~85% of them span 22–6.5 kyr ago. We recalibrated all radiocarbon dates with the IntCal04 calibration (Supplementary Information) and converted proxy units to temperature using the reservoir corrections and proxy calibrations suggested in the original publications. An exception to this was the alkenone records, which were recalibrated with a global core-top calibration<sup>41</sup>. The data were projected onto a 5° × 5° grid, linearly interpolated to 100-yr resolution and combined as area-weighted averages. We used Monte Carlo simulations to quantify pooled uncertainties in the age models and proxy temperatures, although we do not account for analytical uncertainties or uncertainties related to lack of global coverage and spatial bias in the data set. In particular, the records are strongly biased towards ocean margins where high sedimentation rates facilitate the development of high-resolution records. Given these issues, we focus on the temporal evolution of temperature through the deglaciation rather than on its amplitude of change. The global temperature stack is not particularly sensitive to interpolation resolution, areal weighting, the number of proxy records, radiocarbon calibration, infilling of missing data or proxy type. Details on the experimental design of the transient model simulations can be found in ref. 25. The temperature stacks and proxy data set are available in Supplementary Information.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 16 September 2011; accepted 1 February 2012.**

- Hays, J. D., Imbrie, J. & Shackleton, N. J. Variations in the Earth's orbit: pacemaker of the ice ages. *Science* **194**, 1121–1132 (1976).
- Delmas, R. J., Ascencio, J. M. & Legrand, M. Polar ice evidence that atmospheric CO<sub>2</sub> 20,000 yr BP was 50% of present. *Nature* **284**, 155–157 (1980).
- Neftel, A., Oeschger, H., Schwander, J., Stauffer, B. & Zumbunn, R. Ice core sample measurements give atmospheric CO<sub>2</sub> content during the past 40,000 yr. *Nature* **295**, 220–223 (1982).
- Lüthi, D. *et al.* High-resolution carbon dioxide concentration record 650,000–800,000 years before present. *Nature* **453**, 379–382 (2008).
- Shackleton, N. J. The 100,000 year ice-age cycle identified and found to lag temperature, carbon dioxide and orbital eccentricity. *Science* **289**, 1897–1902 (2000).
- Imbrie, J. *et al.* On the structure and origin of major glaciation cycles. 2. The 100,000-year cycle. *Paleoceanography* **8**, 699–735 (1993).
- Alley, R. B. & Clark, P. U. The deglaciation of the northern hemisphere: a global perspective. *Annu. Rev. Earth Planet. Sci.* **27**, 149–182 (1999).
- Toggweiler, J. R. & Lea, D. W. Temperature differences between the hemispheres and ice age climate variability. *Paleoceanography* **25**, PA2212 (2010).
- Weaver, A. J., Eby, M., Fanning, A. F. & Wiebe, E. C. Simulated influence of carbon dioxide, orbital forcing and ice sheets on the climate of the Last Glacial Maximum. *Nature* **394**, 847–853 (1998).
- Schneider von Deimling, T., Held, H., Ganopolski, A. & Rahmstorf, S. Climate sensitivity estimated from ensemble simulations of glacial climate. *Clim. Dyn.* **27**, 149–163 (2006).
- Mix, A. C., Ruddiman, W. F. & McIntyre, A. Late Quaternary paleoceanography of the tropical Atlantic, 1: spatial variability of annual mean sea-surface temperatures, 0–20,000 years B.P. *Paleoceanography* **1**, 43–66 (1986).
- Monnin, E. *et al.* Atmospheric CO<sub>2</sub> concentrations over the last glacial termination. *Science* **291**, 112–114 (2001).
- Lemieux-Dudon, B. *et al.* Consistent dating for Antarctic and Greenland ice cores. *Quat. Sci. Rev.* **29**, 8–20 (2010).



14. Fischer, H., Wahlen, M., Smith, J., Mastoianni, D. & Deck, B. Ice core records of atmospheric CO<sub>2</sub> around the last three glacial terminations. *Science* **283**, 1712–1714 (1999).
15. Hansen, J. *et al.* Climate response times: dependence on climate sensitivity and ocean mixing. *Science* **229**, 857–859 (1985).
16. Manabe, S. & Broccoli, A. J. The influence of continental ice sheets on the climate of an ice age. *J. Geophys. Res.* **90**, 2167–2190 (1985).
17. Broccoli, A. J. Tropical cooling at the Last Glacial Maximum: an atmosphere-mixed layer ocean model simulation. *J. Clim.* **13**, 951–976 (2000).
18. Chiang, J. C. H. & Bitz, C. M. Influence of high latitude ice cover on the marine Intertropical Convergence Zone. *Clim. Dyn.* **25**, 477–496 (2005).
19. Jansen, E. *et al.* in *Climate Change 2007: The Physical Science Basis* (eds Solomon, S. *et al.*) 433–497 (Cambridge Univ. Press, 2007).
20. Clark, P. U. *et al.* Global climate evolution during the last deglaciation. *Proc. Natl Acad. Sci. USA* advance online publication doi:10.1073/pnas.1116619109 (13 February 2012).
21. Blunier, T. *et al.* Synchronization of ice core records via atmospheric gases. *Clim. Past* **3**, 325–330 (2007).
22. Crowley, T. J. North Atlantic Deep Water cools the Southern Hemisphere. *Paleoceanography* **7**, 489–497 (1992).
23. Stocker, T. F. & Johnsen, S. J. A minimum thermodynamic model for the bipolar seesaw. *Paleoceanography* **18**, 1087 (2003).
24. McManus, J. F., Francois, R., Gherardi, J.-M., Keigwin, L. D. & Brown-Leger, S. Collapse and rapid resumption of Atlantic meridional circulation linked to deglacial climate changes. *Nature* **428**, 834–837 (2004).
25. Liu, Z. *et al.* Transient simulation of last deglaciation with a new mechanism for Bölling-Allerød warming. *Science* **325**, 310–314 (2009).
26. Clark, P. U. *et al.* The Last Glacial Maximum. *Science* **325**, 710–714 (2009).
27. Schmittner, A. *et al.* Climate sensitivity estimated from temperature reconstructions of the Last Glacial Maximum. *Science* **334**, 1385–1388 (2011).
28. Alley, R. B., Brook, E. J. & Anandakrishnan, S. A northern lead in the orbital band: north-south phasing of Ice-Age events. *Quat. Sci. Rev.* **21**, 431–441 (2002).
29. Yokoyama, Y., Lambeck, K., De Deckker, P., Johnston, P. & Fifield, L. K. Timing of the Last Glacial Maximum from observed sea-level minima. *Nature* **406**, 713–716 (2000).
30. Clark, P. U., McCabe, A. M., Mix, A. C. & Weaver, A. J. Rapid rise of sea level 19,000 years ago and its global implications. *Science* **304**, 1141–1144 (2004).
31. Marchitto, T. M., Lehman, S. J., Ortiz, J. D., Fluckiger, J. & van Geen, A. Marine radiocarbon evidence for the mechanism of deglacial atmospheric CO<sub>2</sub> rise. *Science* **316**, 1456–1459 (2007).
32. Skinner, L. C., Fallon, S., Waelbroeck, C., Michel, E. & Barker, S. Ventilation of the deep Southern Ocean and deglacial CO<sub>2</sub> rise. *Science* **328**, 1147–1151 (2010).
33. Stephens, B. B. & Keeling, R. F. The influence of Antarctic sea ice on glacial-interglacial CO<sub>2</sub> variations. *Nature* **404**, 171–174 (2000).
34. Toggweiler, J. R., Russell, J. L. & Carson, S. R. Midlatitude westerlies, atmospheric CO<sub>2</sub>, and climate change during the ice ages. *Paleoceanography* **21**, PA2005 (2006).
35. Schmittner, A. & Galbraith, E. D. Glacial greenhouse-gas fluctuations controlled by ocean circulation changes. *Nature* **456**, 373–376 (2008).
36. Barker, S. *et al.* Interhemispheric Atlantic seesaw response during the last deglaciation. *Nature* **457**, 1097–1102 (2009).
37. Schmittner, A., Saenko, O. & Weaver, A. J. Coupling of the hemispheres in observations and simulations of glacial climate change. *Quat. Sci. Rev.* **22**, 659–671 (2003).
38. Anderson, R. F. *et al.* Wind-driven upwelling in the Southern Ocean and the deglacial rise in atmospheric CO<sub>2</sub>. *Science* **323**, 1443–1448 (2009).
39. Stott, L., Timmermann, A. & Thunell, R. Southern hemisphere and deep-sea warming led deglacial atmospheric CO<sub>2</sub> rise and tropical warming. *Science* **318**, 435–438 (2007).
40. Huybers, P. & Denton, G. Antarctic temperature at orbital timescales controlled by local summer duration. *Nature Geosci.* **1**, 787–792 (2008).
41. Müller, P. J., Kirst, G., Ruhland, G., von Storch, I. & Rosell-Mele, A. Calibration of the alkenone paleotemperature index U<sub>37<sup>K</sup> based on core-tops from the eastern South Atlantic and the global ocean (60°N–60°S). *Geochim. Cosmochim. Acta* **62**, 1757–1772 (1998).</sub>
42. Pedro, J. B. *et al.* The last deglaciation: timing the bipolar seesaw. *Clim. Past Discuss.* **7**, 397–430 (2011).
43. Dyke, A. S. in *Quaternary Glaciations: Extent and Chronology* Vol. 2b (eds Ehlers, J. & Gibbard, P. L.) 373–424 (Elsevier, 2004).
44. Laskar, J. *et al.* A long term numerical solution for the insolation quantities of the Earth. *Astron. Astrophys.* **428**, 261–285 (2004).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** Discussions with numerous people, including E. J. Brook, A. E. Carlson, N. G. Pisias and J. Shaman, contributed to this research. We acknowledge the palaeoclimate community for generating the proxy data sets used here. In particular, we thank S. Barker, T. Barrows, E. Calvo, J. Kaiser, A. Koutavas, Y. Kubota, V. Peck, C. Pelejero, J.-R. Petit, J. Sachs, E. Schefuß, J. Tierney and G. Wei for providing proxy data, and R. Gyllencreutz and J. Mangerud for providing unpublished results of the DATED Project on the retreat history of the Eurasian ice sheets. The NOAA NGDC and PANGAEA databases were also essential to this work. This research used resources of the Oak Ridge Leadership Computing Facility, located in the National Center for Computational Sciences at Oak Ridge National Laboratory, which is supported by the Office of Science of the Department of Energy under contract no. DE-AC05-00OR22725. NCAR is sponsored by the NSF. J.D.S. is supported by a NOAA Climate and Global Change Postdoctoral Fellowship. This research was supported by the NSF Paleoclimate Program for the Paleovar Project through grant AGS-0602395.

**Author Contributions** J.D.S. designed the study, synthesized and analysed data, and wrote the manuscript with P.U.C. F.H., Z.L. and B.O.-B. did the transient modelling. S.A.M. and A.C.M. contributed to data analysis. A.S. helped interpret AMOC–CO<sub>2</sub> linkages. E.B. provided data and discussion on the radiocarbon calibration. All authors discussed the results and provided input on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to J.D.S. ([shakun@fas.harvard.edu](mailto:shakun@fas.harvard.edu)).

## METHODS

**Age control.** All radiocarbon dates were recalibrated using Calib 6.0.1 with the IntCal04 calibration and the reservoir corrections suggested in the original publications. Age models based on tuning were left unchanged from the original publications. We used the GICC05 timescale for NGRIP and GRIP, the timescale of ref. 13 for EDML and Dome C, and glaciological age models for the Dome F and Vostok ice cores.

**Proxy temperatures.** We converted proxy units to temperature for all alkenone, Mg/Ca and TEX<sub>86</sub> records using the calibrations suggested by the original authors for Mg/Ca and TEX<sub>86</sub> and the global core-top calibration for alkenone records<sup>41</sup>. We used the published temperature reconstructions for Antarctic ice-core, pollen, microfossil assemblage and MBT/CBT records and the GISP2 borehole calibration for the Greenland ice cores<sup>45</sup>. Missing data values near the beginning and end of the ~15% of records not spanning the entire study interval were infilled using the method of regularized expectation maximization<sup>46</sup>.

**Stacking.** The proxy data were projected onto a 5° × 5° grid, linearly interpolated to 100-yr resolution and combined as area-weighted averages. We do not otherwise account for spatial biases in the dataset or lack of global coverage.

**Uncertainty analysis.** There are two main sources of uncertainty in the proxy records: age models and temperature calibration. We used a Monte Carlo approach to generate 1,000 realizations of each proxy temperature record after perturbing the records with chronological and temperature errors. These perturbed records were then averaged to yield 1,000 realizations of the global and hemispheric temperature stacks. The error bars on the temperature stacks represent the standard deviations of these 1,000 realizations, which provide an estimate of the propagated uncertainty due to uncertainties in the individual proxy records. A similar approach was applied to the transient model output to develop the modelled temperature stacks and error bars. We developed continuous uncertainty estimates for radiocarbon-based chronologies, taking into account radiocarbon date errors as well as interpolation uncertainty between dates using a random walk model<sup>47</sup>. Age-model uncertainties for tuned records, the Dome F and Vostok ice cores, and regional temperature reconstructions for Beringia<sup>48</sup> were assumed to be 2% (1σ). We used the Dome C and EDML ice-core age-model uncertainties<sup>13</sup> and GICC05 maximum counting errors<sup>49,50</sup> as 2σ uncertainties for the NGRIP and GRIP ice cores as suggested in ref. 50. We used the following 1σ temperature calibration uncertainties: alkenones,  $T = (U_{37}^K - 0.044 \pm 0.016) / (0.033 \pm 0.001)$  (ref. 41); Mg/Ca =  $\pm 0.02B_{exp}(\pm 0.003AT)$ , where A and B are constants (ref. 51); TEX<sub>86</sub>,  $\pm 1.7^\circ\text{C}$  (ref. 52); ice cores,  $\pm 10\%$  (ref. 53); pollen, microfossil assemblages and MBT/CBT,  $\pm 1.5^\circ\text{C}$ . We did not account for analytical uncertainties in proxy measurements. Chronological errors in the Monte Carlo simulations were temporally autocorrelated but were random in space (but this does not account for systematic errors among the proxy records due to uncertainties in the radiocarbon calibration), whereas temperature errors were assumed to be random in space and time. We note that our study is concerned with temperature anomalies and is thus sensitive to relative but not absolute temperature errors in a proxy record. See Supplementary Information for more details and examples. Age-model uncertainties for the Antarctic Dome C CO<sub>2</sub> record related to methane synchronization to Greenland were estimated on the basis of the combined uncertainties associated with Greenland layer counting, Greenland ice-age/gas-age differences and methane tuning to Antarctica. These uncertainties were used to generate 1,000 realizations of the CO<sub>2</sub> record, which together with the 1,000 temperature stack realizations yield the 1,000 temperature–concentration lead–lag estimates shown in Fig. 2b.

**Robustness of results.** We evaluated how well the proxy sites represent the globe by subsampling the twentieth-century instrumental temperature record and our transient modelling output of the deglaciation at the 80 proxy sites. Both approaches suggested that the mean of the proxy sites approximates the global mean fairly well. We recalculated the stack after interpolating the records to

500-yr resolution but this did not change the time series or its uncertainty. Differences in areal weighting affect the glacial–interglacial amplitude of the stack but have little impact on its structure. Jackknifing suggests that the stack is not particularly sensitive to the number of records used. A leave-one-out proxy jackknifing approach suggests that the correlation ( $r^2 = 0.90\text{--}0.95$ ) and lead–lag relation (300–600-yr temperature lag) between global temperature and CO<sub>2</sub> concentration are not sensitive to proxy type. Statistical infilling of missing data values has negligible impact on the results. Although we here use the IntCal04 radiocarbon calibration, we tested the sensitivity of our results to this choice by recalibrating radiocarbon dates using the IntCal09 calibration. This makes the global stack up to 350 yr older during the Heinrich 1 interval, and shifts the overall phase lag relative to CO<sub>2</sub> concentration from  $460 \pm 340$  to  $350 \pm 340$  yr. We consider the IntCal04 calibration to be more accurate for the reasons discussed in Supplementary Information. Lag correlations suggest that Antarctic temperature led CO<sub>2</sub> concentration slightly throughout the deglaciation, whereas global temperature led CO<sub>2</sub> concentration at the onset of deglaciation but lagged behind it thereafter. The lead–lag relation between CO<sub>2</sub> concentration and the global temperature stack is not significantly changed by detrending the time series to remove the deglacial ramp in each quantity. The significance levels of the correlations between global temperature and CO<sub>2</sub> concentration and between Pa/Th and ΔT were determined by calculating effective sample sizes of these highly autocorrelated time series (CO<sub>2</sub> concentration, 4.3; global temperature, 4.1; Pa/Th, 5.4; ΔT, 6.2; Atlantic ΔT, 5.5) using equation 6.26 of ref. 54. See Supplementary Information for more discussion of these tests.

**Model freshwater forcing.** Whereas the forcing from insolation, greenhouse gases and ice sheets during the deglaciation are fairly well constrained, freshwater forcing is comparatively uncertain. Several model freshwater schemes were tested, and the final run was based on the meltwater scenario (Supplementary Fig. 30) that produced North Atlantic climate variability in best agreement with proxy reconstructions. The raw modelled AMOC time series (Fig. 4f, thin lines) were effectively smoothed with a Monte Carlo approach similar to the one used to develop the modelled temperature stacks (Fig. 4e), to facilitate direct comparison of the two. More specifically, the smoothed AMOC time series (Fig. 4f, bold lines) are the means of 1,000 realizations of the raw AMOC time series generated by perturbing them with 300-yr (1σ) age-model errors.

45. Cuffey, K. M. & Clow, G. D. Temperature, accumulation, and ice sheet elevation in central Greenland through the last deglacial transition. *J. Geophys. Res.* **102**, 26383–26396 (1997).
46. Schneider, T. Analysis of incomplete climate data: estimation of mean values and covariance matrices and imputation of missing values. *J. Clim.* **14**, 853–871 (2001).
47. Huybers, P. & Wunsch, C. A depth-derived Pleistocene age model: uncertainty estimates, sedimentation variability, and nonlinear climate change. *Paleoceanography* **19**, PA1028 (2004).
48. Viau, A. E., Gajewski, K., Sawada, M. C. & Bunbury, J. Low- and high-frequency climate variability in eastern Beringia during the past 25 000 years. *Can. J. Earth Sci.* **45**, 1435–1453 (2008).
49. Rasmussen, S. O. *et al.* Synchronization of the NGRIP, GRIP, and GISP2 ice cores across MIS 2 and palaeoclimatic implications. *Quat. Sci. Rev.* **27**, 18–28 (2008).
50. Svensson, A. *et al.* A 60000 year Greenland stratigraphic ice core chronology. *Clim. Past* **4**, 47–57 (2008).
51. Anand, P., Elderfield, H. & Conte, M. H. Calibration of Mg/Ca thermometry in planktonic foraminifera from a sediment trap time series. *Paleoceanography* **18**, 1050 (2003).
52. Kim, J. H., Schouten, S., Hopmans, E. C., Donner, B. & Damste, J. S. S. Global sediment core-top calibration of the TEX<sub>86</sub> paleothermometer in the ocean. *Geochim. Cosmochim. Acta* **72**, 1154–1173 (2008).
53. Jouzel, J. *et al.* Magnitude of isotope/temperature scaling for interpretation of central Antarctic ice cores. *J. Geophys. Res.* **108**, 4361 (2003).
54. von Storch, H. & Zwiers, F. W. *Statistical Analysis in Climate Research* 115 (Cambridge Univ. Press, 1999).

# The genomic basis of adaptive evolution in threespine sticklebacks

Felicity C. Jones<sup>1\*</sup>, Manfred G. Grabherr<sup>2,3\*</sup>, Yingguang Frank Chan<sup>1†\*</sup>, Pamela Russell<sup>2\*</sup>, Evan Mauceli<sup>2†</sup>, Jeremy Johnson<sup>2</sup>, Ross Swofford<sup>2</sup>, Mono Pirun<sup>2†</sup>, Michael C. Zody<sup>2</sup>, Simon White<sup>4</sup>, Ewan Birney<sup>5</sup>, Stephen Searle<sup>4</sup>, Jeremy Schmutz<sup>6</sup>, Jane Grimwood<sup>6</sup>, Mark C. Dickson<sup>6</sup>, Richard M. Myers<sup>6</sup>, Craig T. Miller<sup>1†</sup>, Brian R. Summers<sup>1</sup>, Anne K. Knecht<sup>1</sup>, Shannon D. Brady<sup>1</sup>, Haili Zhang<sup>1</sup>, Alex A. Pollen<sup>1</sup>, Timothy Howes<sup>1</sup>, Chris Amemiya<sup>7</sup>, Broad Institute Genome Sequencing Platform & Whole Genome Assembly Team<sup>‡</sup>, Eric S. Lander<sup>2</sup>, Federica Di Palma<sup>2</sup>, Kerstin Lindblad-Toh<sup>2,3</sup> & David M. Kingsley<sup>1,8</sup>

Marine stickleback fish have colonized and adapted to thousands of streams and lakes formed since the last ice age, providing an exceptional opportunity to characterize genomic mechanisms underlying repeated ecological adaptation in nature. Here we develop a high-quality reference genome assembly for threespine sticklebacks. By sequencing the genomes of twenty additional individuals from a global set of marine and freshwater populations, we identify a genome-wide set of loci that are consistently associated with marine–freshwater divergence. Our results indicate that reuse of globally shared standing genetic variation, including chromosomal inversions, has an important role in repeated evolution of distinct marine and freshwater sticklebacks, and in the maintenance of divergent ecotypes during early stages of reproductive isolation. Both coding and regulatory changes occur in the set of loci underlying marine–freshwater evolution, but regulatory changes appear to predominate in this well known example of repeated adaptive evolution in nature.

The genetic and molecular basis of adaptive evolution is still largely unknown. Some researchers have championed a pre-eminent role for regulatory changes during evolution of adaptive phenotypes, because such changes may avoid pleiotropic consequences of protein-coding alterations<sup>1–3</sup>. Others have catalogued known phenotypic differences caused by protein-coding changes and have questioned whether sufficient case histories exist to estimate the relative frequency of regulatory and coding changes during adaptive evolution<sup>4</sup>. Despite progress on individual traits<sup>5</sup>, it has been difficult to accumulate enough examples in any particular group to obtain an overall picture of molecular mechanisms underlying evolutionary change, particularly for clearly adaptive phenotypes in wild organisms.

Threespine sticklebacks offer a powerful system for studying the molecular basis of adaptive evolution in vertebrates. After the retreat of Pleistocene glaciers, marine sticklebacks colonized and adapted to many newly formed freshwater habitats, evolving repeated changes in body shape, skeletal armour, trophic specializations, pigmentation, salt handling, life history and mating preferences<sup>6,7</sup>. Recurrent evolution of similar phenotypes in similar environments indicates that these traits evolve by natural selection<sup>8</sup>. Distinctive marine and freshwater forms can still hybridize, making it possible to map the genetic basis of individual traits, and identify particular genes underlying armour, pelvic and pigmentation evolution<sup>9–12</sup>. At two of these key loci, distinctive haplotypes were found to be reused when similar phenotypes evolve in different populations<sup>11,12</sup>, a pattern that was later found at additional loci<sup>13,14</sup>. Ongoing gene flow between marine and freshwater forms occurs along coastal rivers<sup>15,16</sup>, making it possible to spread

adaptive alleles among populations, and homogenizing neutral genomic regions<sup>17</sup>. Here we use signatures of allele sharing to identify a genome-wide set of adaptive loci consistently associated with recurrent marine–freshwater evolution.

## Generation of reference genome assembly

To facilitate studies of stickleback evolution, we first generated a reference genome assembly from a homogametic (female) freshwater stickleback (*Gasterosteus aculeatus*) from Bear Paw Lake, Alaska. The sequenced individual was partially inbred and retained heterozygosity at approximately 1 per 700 base pairs (bp). The assembly, gasAcul.0, was generated with 9.0× coverage in Sanger sequence data (ABI3730), and has a length-weighted median (N50) contig size of 83.2 kilobases (kb), a length-weighted median (N50) scaffold size of 10.8 megabases (Mb) and a total gapped size of 463 Mb, close to previous estimates of 530 Mb (ref. 18). The 113 largest scaffolds (86.9%, 400.4 Mb) were anchored to stickleback linkage groups in an F<sub>2</sub> marine × freshwater intercross, whereas 60.7 Mb in 1,812 smaller scaffolds (N50 = 0.3 Mb) remain unanchored. Use of a single partially inbred individual, construction and assembly of a range of genomic library sizes, and the relatively low repeat and duplication content of the stickleback genome have produced a highly contiguous anchored genome assembly with contig and scaffold sizes much larger than other published teleosts<sup>19–22</sup> (Supplementary Table 1).

The stickleback sequence was annotated using the Ensembl pipeline, which predicted 20,787 protein-coding and 1,617 RNA genes (Supplementary Table 2). Of the protein-coding genes, 7,614 showed

<sup>1</sup>Department of Developmental Biology, Beckman Center B300, Stanford University School of Medicine, Stanford California 94305, USA. <sup>2</sup>Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge Massachusetts 02142, USA. <sup>3</sup>Science for Life Laboratory Uppsala, Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala 751 23, Sweden. <sup>4</sup>Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK. <sup>5</sup>European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. <sup>6</sup>HudsonAlpha Institute for Biotechnology, 601 Genome Way, Huntsville, Alabama 35806, USA. <sup>7</sup>Department of Molecular Genetics, Benaroya Research Institute at Virginia Mason, 1201 Ninth Avenue, Seattle Washington 98101, USA. <sup>8</sup>Howard Hughes Medical Institute, Stanford University, Stanford, California 94305, USA. <sup>†</sup>Present addresses: Max Planck Institute for Evolutionary Biology, August-Thienemann-Str. 2, Plön 24306, Germany (Y.F.C.); Children's Hospital Boston, Genetic Diagnostic Lab, 300 Longwood Avenue, Boston, Massachusetts 02115, USA (E.M.); Bioinformatics Core, Zuckerman Research Center, New York, New York 10065, USA (M.P.); Department of Molecular & Cell Biology, 142 LSA 3200, University of California, Berkeley, California 94720, USA (C.T.M.).

\*These authors contributed equally to this work.

‡Lists of authors and affiliations appear at the end of the paper.



one-to-one orthology with mammals and an additional 7,192 showed one-to-one orthology among fishes. The other 5,981 genes showed complex orthology relationships, including some lineage-specific gene expansions that contribute to stickleback adaptations (for example, a duplicated mucin family encoding glue proteins used for male nest building<sup>23</sup>). A total of 13.4% of the stickleback genome appeared to be under evolutionary constraint when compared with other fishes using PhastCons<sup>24</sup>. The conserved portion was roughly equally divided between protein-coding and non-coding sequences, with ~71% of the latter shared with mammals and ~29% representing fish-specific conserved sequences (Supplementary Table 3).

### Sequencing additional population pairs

To search for loci underlying repeated evolution in sticklebacks, we first identified populations showing characteristic marine and freshwater morphology (Fig. 1a, Supplementary Fig. 1 and Supplementary Table 4). Repeated adaptation to divergent marine and freshwater environments resulted in marked correlated changes in body shape, length, depth, fin position, spine length, eye size and armour plate number (Fig. 1b). Because quantitative trait loci (QTL) controlling these traits map to many different chromosomes<sup>12,25–30</sup>, this morphological screen should identify populations differing in a genome-wide range of adaptive loci underlying marine–freshwater differences.

From the distinct morphological clusters of marine and freshwater fish, we selected multiple marine–freshwater pairs, from both Pacific and Atlantic populations, including individuals from opposite ends of rivers with marine–freshwater hybrid zones<sup>16</sup> (21 fish in total, including the reference genome individual). The sampling strategy should

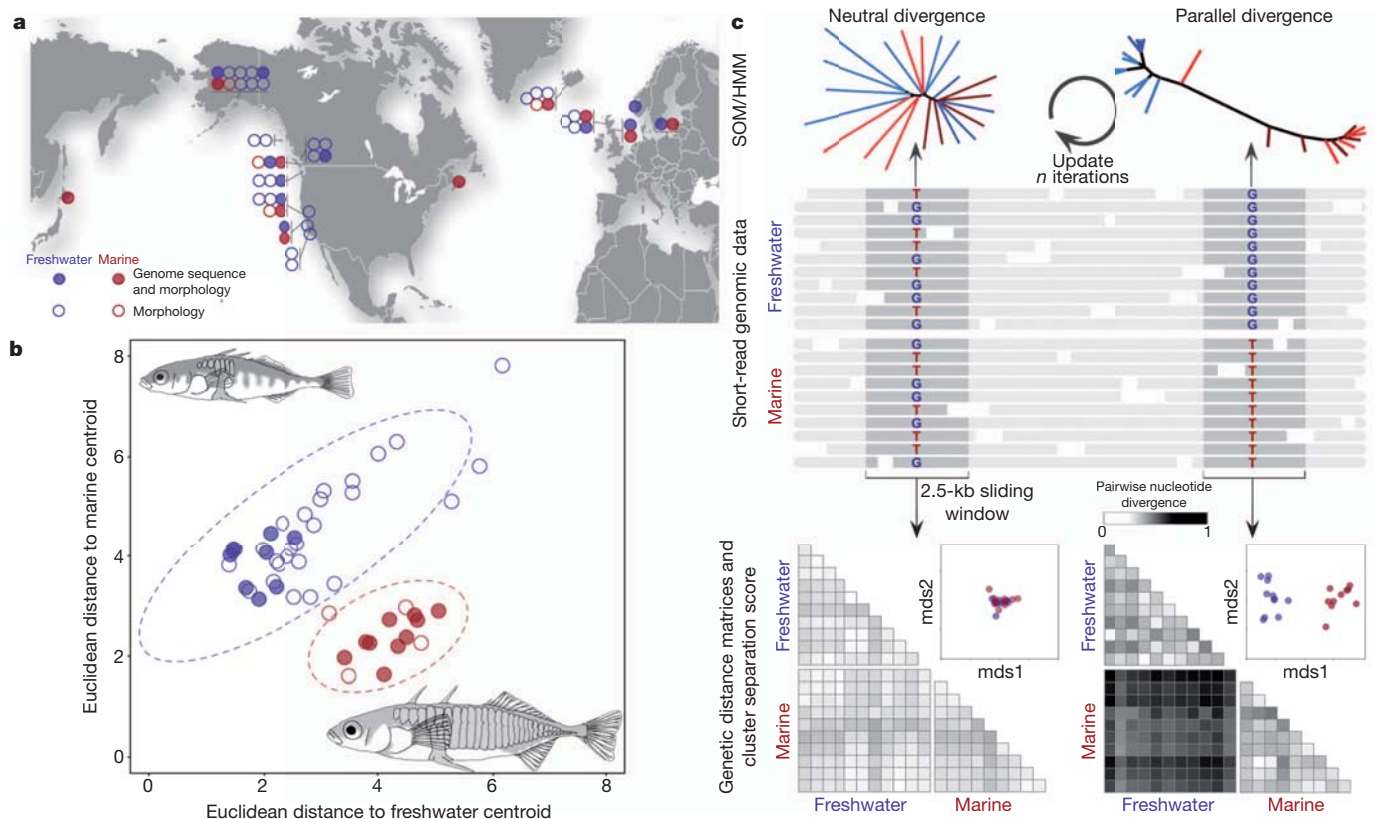
minimize geographic bias in the data set, while maximizing the chance for local exchange of neutral regions of the genome.

We generated  $2.3\times$  average coverage per individual using Illumina sequencing (Supplementary Table 5 and Supplementary Information). To identify single nucleotide polymorphisms (SNPs), we pooled data from all fish and identified positions where at least four reads support a variant allele. This criterion identified 5,897,368 candidate SNPs (Supplementary Table 6), with most being true positives based on experimental validation ( $n = 48$  tested, 82.6% confirmed; Supplementary Information).

### Genome-wide survey of parallel evolution

Previous studies have shown that repeated armour evolution in sticklebacks occurs through ancient variants at the *EDA* locus, which are reused in multiple freshwater populations<sup>11</sup> and are subject to strong selection<sup>31</sup>. To identify loci where alleles have similarly been used repeatedly during adaptive divergence of marine and freshwater fish, we used two methods to look for regions where sequences of most freshwater fish were similar to each other, but differed from sequences typically found in marine populations. Note that this pattern will not identify adaptive variants that are unique to individual freshwater populations, but instead focuses on variants with striking evidence of biological replication across populations.

First, we developed a self-organizing map-based iterative Hidden Markov Model (SOM/HMM) to identify the 20 most common patterns of genetic relationships ('trees') among the 21 individuals. Genomic regions were assigned to pattern types on the basis of likelihood, with boundaries defined using HMM transitions. This method iteratively



**Figure 1 | Genome scans for parallel marine–freshwater divergence.**

**a**, Marine (red) and freshwater (blue) stickleback populations were surveyed from diverse locations. **b**, Morphometric analysis was used to select individuals for re-sequencing. The 20 chosen individuals are from multiple geographically proximate pairs of populations with typical marine and freshwater morphology (solid symbols). Points, population mean morphologies; ellipses, 95% confidence intervals for ecotypes. **c**, Genomes were analysed using SOM/HMM

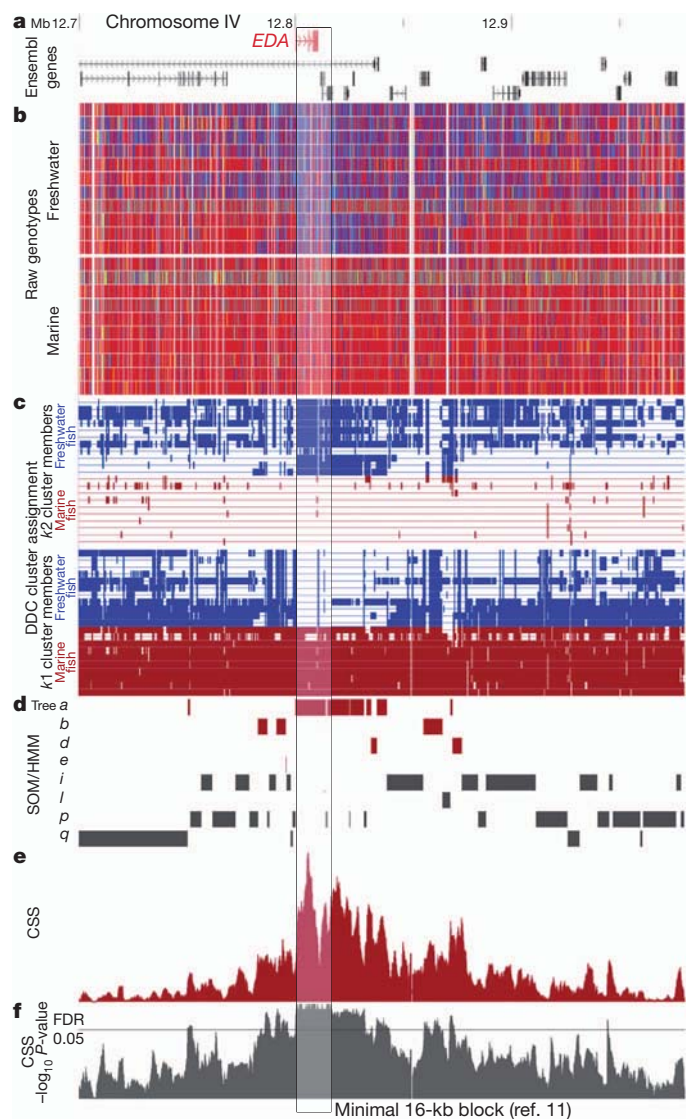
(top) and CSS (bottom) methods to identify parallel marine–freshwater divergent regions. Across most of the genome, the dominant patterns reflect neutral divergence or geographic structure. In contrast, <0.5% of the genome shows clustering by ecotype, a pattern characteristic of divergent marine and freshwater adaptation via parallel reuse of standing genetic variation<sup>11,12</sup>. Mds1 and mds2 represent the first and second major axes of variation extracted from pairwise genetic distance matrices using multidimensional scaling.

models recurring phylogenetic patterns on a local genomic basis with increasing resolution (Fig. 1c and Supplementary Information). Most of the genome was assigned to trees describing geographic relationships between populations (for example, distinct Pacific versus Atlantic clades, each containing marine and freshwater fish; Supplementary Table 7 and Supplementary Figs 2 and 3). A total of 215 regions comprising 2,096,101 bp (0.46% of the genome; median size: 4,684 bp) were assigned to one of four trees separating most marine from most freshwater fish (Supplementary Fig. 3, trees a–d). After filtering, the most prevalent marine–freshwater divergent tree identified 90 genomic regions with a median size of 4,266 bp covering 848,691 bp (0.18% of the genome).

Second, we used a genetic distance-based approach (Fig. 1c) based on building  $21 \times 21$  pairwise nucleotide divergence ( $\pi$ ) matrices for each of 877,568 overlapping windows across the genome (2,500 bp, step size: 500 bp). Each distance matrix was used to calculate a marine–freshwater cluster separation score (CSS), quantifying the average distance between marine and freshwater clusters after accounting for variance within ecotypes (Supplementary Information). The score is highly correlated with genetic distance ( $F_{ST}$ ), but provides increased resolution under high divergence (Supplementary Fig. 4). After permutation testing, we recovered 174 marine–freshwater divergent regions, covering a total of 1,214,500 bp (0.26% of the genome; median size: 3,000 bp) at a 5% false discovery rate (FDR), and 84 divergent regions covering 479,500 bp (0.10% of the genome; median: 4,000 bp) at 2% FDR. To assign cluster membership in highly divergent genomic regions, we also used an unguided Bayesian model-based data-driven clustering (DDC; Fig. 2c and Supplementary Information). For each window of the genome, we estimated the most likely number of distinct clusters of fish ( $k = 0$  to 5) and the cluster memberships.

The independent SOM/HMM and CSS approaches both successfully recover the previously described chromosome IV *EDA* locus among the top-scoring marine–freshwater divergent regions (Fig. 2). Notably, the cluster membership assigned by DDC successfully recapitulates the breakpoints of the minimal 16-kb shared freshwater *EDA* haplotype (Fig. 2c) previously defined by a multi-year positional cloning study of the major locus controlling armour plate differences in sticklebacks<sup>11</sup>. Additional regions were identified on the same chromosome with similar marine–freshwater divergence patterns, including regions surrounding the developmental signalling gene *WNT7B* (Supplementary Fig. 5), and a locus involved in hormone and neurotransmitter binding and metabolism (sulphotransferase 4a1, *SULT4A* (ref. 32)). SOM/HMM and CSS defined many other loci that also show globally shared marine–freshwater divergence, including 242 regions identified by either method (0.5% of the genome), and 147 regions identified by both (0.2% of the genome). The median size of recovered regions (<5 kb) approaches the size of individual genes, and often highlights purely intergenic regions, such as the exclusively non-coding region identified between *BANP* and *RAS* on chromosome XIX (Supplementary Fig. 6). The genomic distribution, sizes and overlaps of recovered regions are described in Fig. 3, Supplementary Fig. 7 and Supplementary Table 8, including a list of specific genes identified in top-scoring regions (Supplementary Data 1). Using genotyping assays for SNPs in 11 regions recovered by both SOM/HMM and CSS analyses, we found that 91% of tested regions show significant enrichment of ecotypic alleles in independent marine and freshwater populations (Supplementary Information). These results confirm that our experimental design successfully identifies both known and novel loci consistently associated with parallel evolution of distinct marine and freshwater ecotypes.

Compared to the genome overall, the 242 regions implicated in repeated marine–freshwater evolution show higher gene density (Supplementary Fig. 8,  $P < 4.5 \times 10^{-13}$ ) and higher concentration of conserved non-coding sequences in intergenic regions (Supplementary Fig. 9,  $P < 1.9 \times 10^{-11}$ ), probably reflecting a more complex regulatory architecture<sup>33</sup>. Gene Ontology analysis shows significant

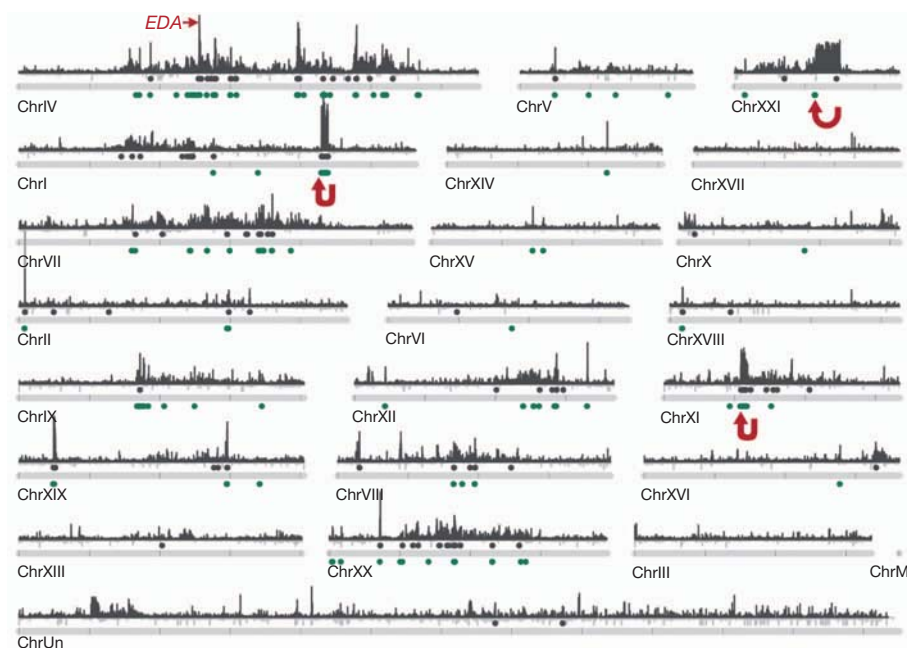


**Figure 2 | Parallel divergence signals at known armour plate locus.**

**a**, Ensembl gene models around *EDA*. **b**, Visual genotypes for sequenced fish (homozygous sites for most frequent allele in marine fish (red); homozygous for alternative allele (blue); heterozygous (yellow), or non-variable/missing/repeat-masked data (white)). **c**, DDC cluster assignments for marine (red) and freshwater populations (blue). Most fish are assigned to cluster *k1*, except in the boxed region, where freshwater fish are assigned to a distinct cluster (*k2*). **d**, SOM/HMM analysis supports patterns of divergence with a marine–freshwater-like tree topology in the centre, but not edges, of the window (trees *a–d*). **e**, **f**, Similar support is shown by CSS analysis (**e**) and its associated *P*-value (**f**). The combined analyses define a consensus 16-kb region shared in freshwater fish (vertical shaded box), matching the minimal haplotype known to control repeated low armour evolution in sticklebacks<sup>11</sup>.

enrichment of genes involved in cellular response to signals, behavioural interaction between organisms, amine and fatty acid metabolism, cell–cell junctions and WNT signalling (Supplementary Table 9). Changes in these biological processes, and in the individual genes defined by parallel divergence analysis, probably underlie recurrent differences in morphology, physiology and behaviour previously described in marine and freshwater sticklebacks<sup>7</sup>. For example, the *WNT7B* and *WNT11* family members identified by the genomic survey have previously been implicated in a paracrine signalling pathway that controls kidney collecting tubule length and diameter<sup>34</sup>. Fish living in fresh water produce copious hypotonic urine compared to marine fish<sup>35</sup>, and long-term adaptation to freshwater may select for





**Figure 3 | Genome-wide distribution of marine-freshwater divergence regions.** Whole-genome profiles of SOM/HMM and CSS analyses reveal many loci distributed on multiple chromosomes (plus unlinked scaffolds, here grouped as 'ChrUn'). Extended regions of marine-freshwater divergence on chromosomes I, XI and XXI correspond to inversions (red arrows).

variants in the same developmental signalling pathways that polarize epithelial cell divisions and regulate kidney tubule formation in other animals.

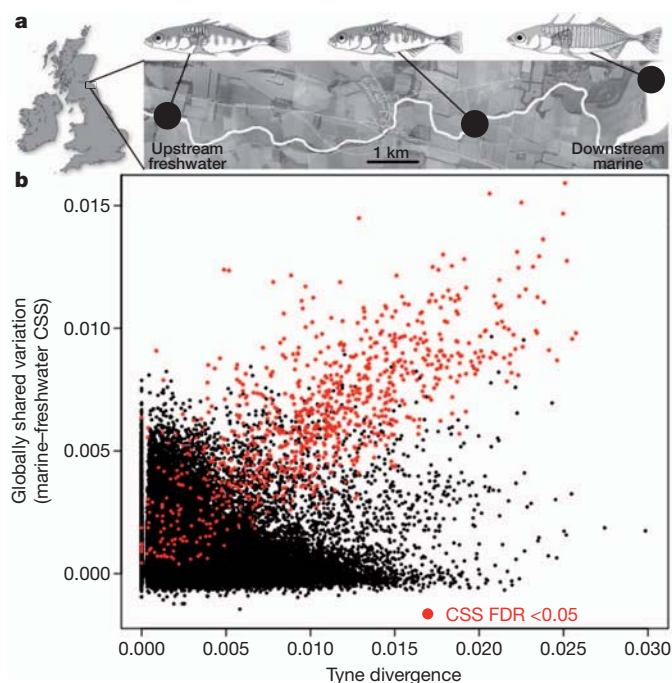
### Extent of parallel reuse in hybrid zones

Although our method identifies regions used repeatedly during stickleback evolution, it does not tell us how prevalent such regions are among all differentiated loci in a particular marine-freshwater species pair. To address this, we analysed patterns of genomic differentiation across a marine-freshwater hybrid zone in River Tyne, Scotland (Fig. 4a). Previous studies show that ecologically mediated postzygotic selection maintains distinct ecotypes in this system, despite hybridization and opportunity for extensive gene flow<sup>16</sup>. Whole-genome sequencing of a pair of marine and freshwater fish from either end of the Tyne hybrid zone identified a set of genomic windows with high divergence. Within the top 0.1% divergent windows, 35.3% contain elevated globally shared marine-freshwater divergence (Fig. 4b and Supplementary Information), indicating an ancient shared origin for many, but not all, loci with highly differentiated alleles in this marine-freshwater species pair. Previous studies have shown that some traits in sticklebacks evolve by independent mutations that vary among populations<sup>10</sup>. The non-globally shared divergent alleles in the Tyne may also represent recent, or locally arising, adaptive variants, although further studies will be required to link such variants to particular traits, or to distinguish them from neutral but highly variable regions of the stickleback genome.

### Marine-freshwater chromosome inversions

When adaptive divergence occurs in hybridizing systems, theory predicts that selection can favour molecular mechanisms that suppress recombination between independent adaptive loci<sup>17</sup>. We observed extended stretches of elevated CSS spanning 442 kb, 412 kb and 1,700 kb on chromosomes I, XI and XXI, respectively (Fig. 3). On the basis of sharp transitions in CSS score and DDC cluster assignments at the boundaries, we hypothesized that chromosomal inversions explain these extended regions. By analysing paired-end sequence reads from a marine large-insert (~220 kb) bacterial

Marine-freshwater divergent regions detected by CSS are shown as grey peaks with grey points above chromosomes indicating regions of significant marine-freshwater divergence (FDR < 0.05). Genomic regions with marine-freshwater-like tree topologies detected by SOM/HMM are shown as green points below chromosomes.

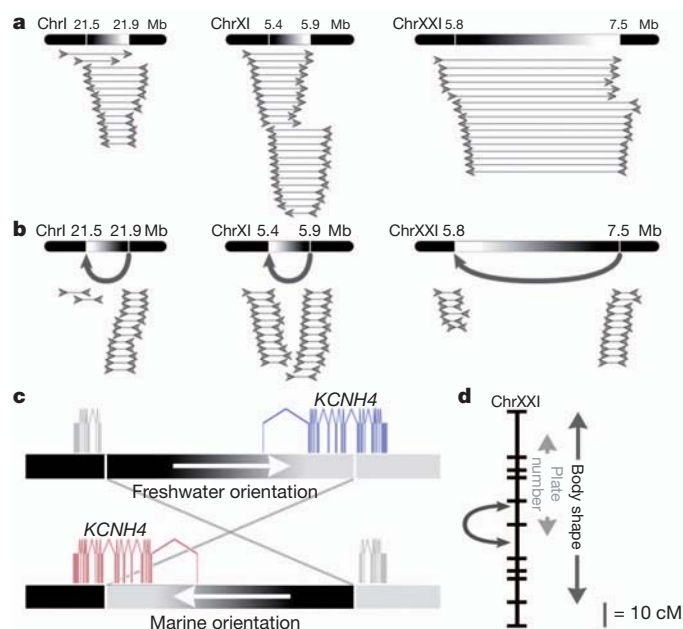


**Figure 4 | How much of local marine-freshwater adaptation occurs by reuse of global variants?** a, Classic marine and freshwater ecotypes are maintained in downstream and upstream locations of the River Tyne, Scotland, despite extensive hybridization at intermediate sites<sup>16</sup>. b, Pairwise sequence comparisons identify many genomic regions that show high divergence between upstream and downstream fish (x axis). Many, but not all, of these regions also show high global marine-freshwater divergence (y axis; red points indicate significant CSS FDR < 0.05), indicating that both global and local variants contribute to formation and reproductive isolation of a marine-freshwater species pair.



artificial chromosome (BAC) library<sup>36</sup>, we identified individual clones with size and orientation anomalies relative to the freshwater reference genome assembly. The only locations with five or more anomalous clones mapped to chromosomes I, XI and XXI, and these anomalies could be resolved by the presence of inverted chromosome segments between the marine fish and the freshwater reference genome (Fig. 5a, b). Sequences flanking the predicted inversion breakpoints contain inverted repeats, consistent with generation of inversions by intra-chromosomal recombination (Supplementary Fig. 10). Notably, repeats flanking the chromosome XI inversion contained alternative 3' exons for the voltage-gated potassium channel gene *KCNH4*. Because *KCNH4* transcription is initiated within the inversion, alternative inversion orientations could generate marine- and freshwater-specific *KCNH4* isoforms (Fig. 5c). Although the functional consequences of such ecotype-specific isoforms remain unknown, *KCNH4* homologues help to maintain resting currents, affect cardiac contractility, and alter performance on cognitive tasks if perturbed in mice<sup>37–39</sup>. Furthermore, QTL for two distinct marine–freshwater divergent traits have previously been mapped to the broad region of the chromosome XXI inversion (Fig. 5d)<sup>27,30</sup>, as expected if inversions help to maintain linkage between different adaptive QTLs<sup>40</sup>.

Importantly, cluster assignment of individual fish by DDC shows that most marine and freshwater populations in the Pacific carry contrasting forms of the inversion regions (Supplementary Table 10). Similar ecotype associations are seen in the Atlantic basin for chromosome I (no exceptions), XI (two exceptions), and to a lesser extent for chromosome XXI (three freshwater exceptions). Genetic markers within the chromosome I and XXI regions are polymorphic in hybrid zones, and show large frequency differences when genotyped in adjacent upstream and downstream fish, confirming that these regions are subject to divergent selection in marine and freshwater habitats (Supplementary Table 10). Our results help to explain the broader



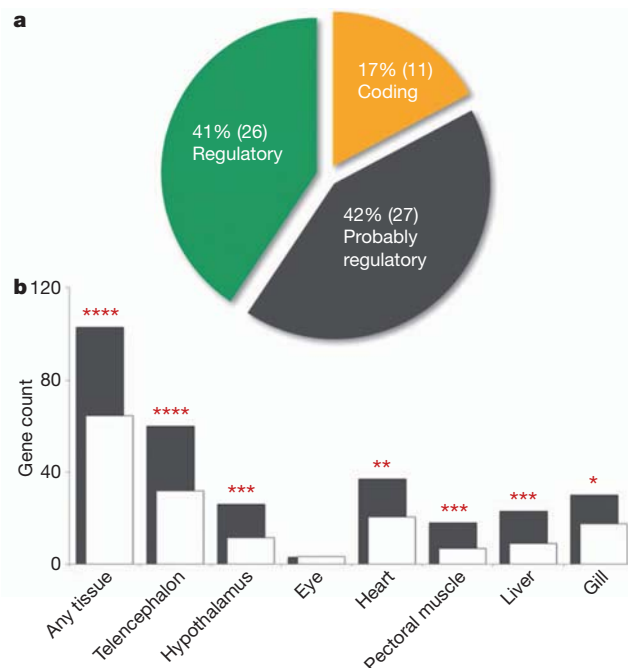
**Figure 5 | Chromosome inversions and marine–freshwater divergence.**

**a**, Multiple marine BAC clones have paired-end reads that place anomalously against the freshwater reference genome (grey arrows below chromosome bars; see Supplementary Methods for BAC names). **b**, Intrachromosomal inversions on chromosomes I, XI and XXI resolve orientation and size anomalies for all marine clones. **c**, The chromosome XI inversion breakpoints map inside the exons of *KCNH4*, a potassium transporter gene. Duplicated 3' exons lead to different transcript orientations and gene products in marine (red gene model) and freshwater fish (blue gene model). **d**, The chromosome XXI inversion occurs in a region with separate QTLs controlling armour plate number and body shape<sup>11,30</sup>, traits that differ between marine and freshwater fish.

patterns of genomic divergence seen in Fig. 3, and add to growing evidence that chromosome inversions are a common genomic mechanism that maintains contrasting ecotypes in hybridizing natural populations<sup>41–44</sup>.

### Proportion of regulatory and coding change

Identification of a genome-wide set of loci used repeatedly in stickleback adaptation provides a rare opportunity to estimate the relative contribution of coding and regulatory changes underlying adaptive evolution in natural populations. To examine this issue, we analysed 64 marine–freshwater divergent regions with the strongest evidence of parallel evolution: those identified by both SOM/HMM and CSS analyses using the strictest significance thresholds (Supplementary Information and Supplementary Data 1), and containing SNPs showing perfect allele–ecotype association between marine and freshwater fish. Many of these 64 regions (41%) mapped entirely to non-coding regions of the genome, and presumably contain regulatory changes (Fig. 6a). A smaller fraction contains protein-coding sequences with consistent non-synonymous substitutions between marine and freshwater fish (17%). Finally, a fraction of regions (43%) include both coding and non-coding sequences (including non-coding RNAs), but lack ecotype-specific amino acid substitutions (Supplementary Data 1). Because all of these regions contain SNPs with perfect allele–ecotype association that do not cause protein-coding changes, they also probably contribute to adaptive divergence by regulatory alterations. The combined data suggest that both coding and regulatory differences contribute to parallel stickleback evolution, with regulatory changes accounting for a much larger proportion of the overall set of loci repeatedly selected during marine–freshwater divergence.



**Figure 6 | Contributions of coding and regulatory changes to parallel marine–freshwater stickleback adaptation.** **a**, A genome-wide set of marine–freshwater divergent loci recovered by both SOM/HMM and CSS analyses includes regions with consistent amino acid substitutions between marine and freshwater ecotypes (yellow sector); regions with no predicted coding sequence (green sector); and regions with both coding and non-coding sequences, but no consistent marine–freshwater amino acid substitutions (grey). **b**, Genome-wide expression analysis shows that marine–freshwater regions identified by SOM/HMM or CSS analyses are enriched for genes showing significant expression differences in 6 out of 7 tissues between marine LTC and freshwater FTC fish (observed, grey bars; expected, white bars; \* $P < 0.01$ , \*\* $P < 0.001$ , \*\*\* $P < 0.0001$ , \*\*\*\* $P < 0.00001$ ), consistent with a role for regulatory changes in marine–freshwater evolution.

To assess further the possible role of gene regulatory evolution in stickleback evolution, we constructed whole-genome expression arrays to compare levels of gene expression in tissues from Little Campbell River (LITC) marine and Fish Trap Creek (FTC) freshwater fish. Of 12,594 informative genes across the genome, 2,817 showed significant expression differences between ecotypes. Genes with marine–freshwater expression differences were significantly more likely to occur in or near the adaptive regions recovered by SOM/HMM or CSS analysis (Fig. 6b,  $P < 7.1 \times 10^{-8}$ ). Although expression differences can be due to either *cis*- or *trans*-acting changes, the expression data are consistent with an important role of regulatory changes during parallel evolution of marine and freshwater sticklebacks.

## Discussion

Progress in genetic mapping and positional cloning approaches has recently made it possible to identify a few individual genes and mutations that contribute to phenotypic differences between stickleback populations<sup>10–12,25</sup>. Despite this progress, identifying many such examples using genetic linkage mapping alone would require years of additional effort. Fortunately, the highly replicated nature of stickleback evolution provides clear molecular signatures that can be used to recover many loci consistently associated with parallel marine–freshwater adaptation. The signal resolution of repeatedly used adaptive loci approaches ~5 kb, often identifying single genes or intergenic regions, and offering a significant advantage over the several hundred kilobase candidate intervals typically identified in genetic mapping crosses<sup>11,12</sup>, or the megabase or larger regions identified in previous selection scans of the stickleback genome<sup>13</sup>. The many marine–freshwater divergent loci and gene expression changes identified in the current study will substantially accelerate ongoing searches for the genetic and molecular basis of fitness-related morphological, physiological and behavioural differences between marine and freshwater fish.

In addition, the genome-wide set of divergent regions already provides new insights into evolutionary processes shaping adaptive evolution and ecological speciation. Our results indicate that parallel evolution of marine and freshwater sticklebacks occurs by dynamic reassembly of many ‘islands’ of divergence distributed across many chromosomes. Reassembly by linkage is probably strengthened by inversions that distinguish marine and freshwater ecotypes. Differences in both globally shared and locally restricted genetic variation actively maintained across a hybrid zone provide a snapshot of the genomic architecture and evolutionary processes contributing to the early stages of reproductive isolation. Finally, our data indicate that repeated evolution of marine–freshwater differences depends on both protein-coding and regulatory changes. Regulatory evolution seems to have a particularly prominent role, as indicated by the increased density of conserved non-coding intergenic sequences found near marine–freshwater divergent loci (Supplementary Fig. 9); the substantial fraction of loci mapping entirely to non-coding regions (Fig. 6a); and the significant enrichment of genes with expression differences near key regions used for parallel evolution (Fig. 6b). Mutations causing structural changes in proteins are the most abundant variants recovered in laboratory *Escherichia coli* and yeast evolution experiments<sup>45,46</sup>. They make up 90% of 40 published examples of adaptive changes between closely related taxa<sup>4</sup>, and 63–77% of the known molecular basis of phenotypic traits in domesticated or wild species<sup>5</sup>. The larger fraction of regulatory changes implicated during repeated stickleback evolution may reflect our use of whole-genome rather than candidate gene approaches, stronger selection against loss-of-function and pleiotropic protein-coding changes in natural populations than in laboratory or domesticated organisms<sup>1–3</sup>, or an increasing prevalence of regulatory changes at interspecific compared to intraspecific levels<sup>5,47</sup>, including emerging species such as marine and freshwater sticklebacks.

Although our study has focused on marine–freshwater divergence, freshwater sticklebacks also repeatedly evolve characteristic lake–stream

differences; open-water and bottom-dwelling lake ecotypes; gigantism in particular lakes; and substantial changes in seasonality and life history<sup>6,7,48–50</sup>. Given the considerable fraction of parallel stickleback evolution probably occurring by shared variants (Fig. 4b), sequencing of additional populations should make it possible to identify similarly shared loci contributing to other ecological traits, again using the power of replicated evolution to illuminate both specific and general mechanisms underlying evolutionary change in natural populations.

## METHODS SUMMARY

A reference stickleback genome sequence was assembled from a single female freshwater stickleback (Bear Paw Lake, Alaska), using 9× coverage of paired-end Sanger-sequenced reads from multiple insert size libraries. Scaffolds were assigned to linkage groups in a genetic cross, and annotation was carried out using the Ensembl evidence-based pipeline. Twenty-one fish from independent populations were chosen for short-read sequencing (48× combined coverage) based on morphometric analysis. Patterns of genetic variation were analysed for divergence between marine and freshwater fish, using both a self-organizing map/Hidden Markov Model and a pairwise distance matrix approach (see Supplementary Information). Paired-end reads from a marine BAC library were placed against the reference freshwater genome sequence to identify possible chromosome rearrangements. Sequenom iPLEX genotyping assays were carried out to verify predicted SNPs and divergent marine–freshwater regions. RNA samples were prepared from tissues of marine and freshwater fish born and raised under identical laboratory conditions. Significant expression differences were detected with Agilent microarrays using eBayes (limma R package). GO category enrichments were analysed using GStats (BioConductor 2.7). Additional methods and analyses are provided in Supplementary Information.

Received 17 September 2011; accepted 8 February 2012.

1. Stern, D. L. Perspective: Evolutionary developmental biology and the problem of variation. *Evolution* **54**, 1079–1091 (2000).
2. Carroll, S. B. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**, 25–36 (2008).
3. Wray, G. The evolutionary significance of *cis*-regulatory mutations. *Nature Rev. Genet.* **8**, 206–216 (2007).
4. Hoekstra, H. E. & Coyne, J. A. The locus of evolution: evo devo and the genetics of adaptation. *Evolution* **61**, 995–1016 (2007).
5. Stern, D. L. & Orgogozo, V. The loci of evolution: how predictable is genetic evolution? *Evolution* **62**, 2155–2177 (2008).
6. McKinnon, J. S. & Rundle, H. D. Speciation in nature: the threespine stickleback model systems. *Trends Ecol. Evol.* **17**, 480–488 (2002).
7. Bell, M. A. & Foster, S. A. *The Evolutionary Biology of the Threespine Stickleback* (Oxford Univ. Press, 1994).
8. Endler, J. A. Natural selection in the wild. *Monogr. Popul. Biol.* **21**, 1–336 (1986).
9. Kingsley, D. M. & Peichel, C. L. The molecular genetics of evolutionary change in sticklebacks. In *Biology of the Threespine Stickleback* 41–81 (CRC Press, 2007).
10. Chan, Y. F. *et al.* Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science* **327**, 302–305 (2010).
11. Colosimo, P. F. *et al.* Widespread parallel evolution in sticklebacks by repeated fixation of *Ectodysplasin* alleles. *Science* **307**, 1928–1933 (2005).
12. Miller, C. T. *et al.* *cis*-Regulatory changes in *Kit* ligand expression and parallel evolution of pigmentation in sticklebacks and humans. *Cell* **131**, 1179–1189 (2007).
13. Hohenlohe, P. A. *et al.* Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet.* **6**, e1000862 (2010).
14. Kitano, J. *et al.* Adaptive divergence in the thyroid hormone signaling pathway in the stickleback radiation. *Curr. Biol.* **20**, 2124–2130 (2010).
15. Hagen, D. Isolating mechanisms in threespine sticklebacks (*Gasterosteus*). *J. Fish. Res. Board Can.* **24**, 1637–1692 (1967).
16. Jones, F., Brown, C., Pemberton, J. & Braithwaite, V. Reproductive isolation in a threespine stickleback hybrid zone. *J. Evol. Biol.* **19**, 1531–1544 (2006).
17. Barton, N. H. & Gale, K. S. Genetic analysis of hybrid zones. In *Hybrid Zones and the Evolutionary Process* 13–45 (Oxford Univ. Press, 1993).
18. Vinogradov, A. E. Genome size and GC-percent in vertebrates as determined by flow cytometry: the triangular relationship. *Cytometry* **31**, 100–109 (1998).
19. Aparicio, S. *et al.* Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* **297**, 1301–1310 (2002).
20. Jaillon, O. *et al.* Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* **431**, 946–957 (2004).
21. Kasahara, M. *et al.* The medaka draft genome and insights into vertebrate genome evolution. *Nature* **447**, 714–719 (2007).
22. Star, B. *et al.* The genome sequence of Atlantic cod reveals a unique immune system. *Nature* **477**, 207–210 (2011).
23. Kawahara, R. & Nishida, M. Extensive lineage-specific gene duplication and evolution of the spiggin multi-gene family in stickleback. *BMC Evol. Biol.* **7**, 209 (2007).
24. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).

25. Shapiro, M. D. *et al.* Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* **428**, 717–723 (2004).
26. Peichel, C. L. *et al.* The genetic architecture of divergence between threespine stickleback species. *Nature* **414**, 901–905 (2001).
27. Colosimo, P. F. *et al.* The genetic architecture of parallel armor plate reduction in threespine sticklebacks. *PLoS Biol.* **2**, 635–641 (2004).
28. Cresko, W. A. *et al.* Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. *Proc. Natl Acad. Sci. USA* **101**, 6050–6055 (2004).
29. Kimmel, C. B. *et al.* Evolution and development of facial bone morphology in threespine sticklebacks. *Proc. Natl Acad. Sci. USA* **102**, 5791–5796 (2005).
30. Albert, A. Y. K. *et al.* The genetics of adaptive shape shift in stickleback: pleiotropy and effect size. *Evolution* **62**, 76–85 (2008).
31. Barrett, R. D. H., Rogers, S. M. & Schluter, D. Natural selection on a major armor gene in threespine stickleback. *Science* **322**, 255–257 (2008).
32. Allali-Hassani, A. *et al.* Structural and chemical profiling of the human cytosolic sulfotransferases. *PLoS Biol.* **5**, e97 (2007).
33. Knecht, A. K., Hosemann, K. E. & Kingsley, D. M. Constraints on utilization of the EDA-signaling pathway in threespine stickleback evolution. *Evol. Dev.* **9**, 141–154 (2007).
34. Yu, J. *et al.* A *Wnt7b*-dependent pathway regulates the orientation of epithelial cell division and establishes the cortico-medullary axis of the mammalian kidney. *Development* **136**, 161–171 (2009).
35. Marshall, W. S. & Grosell, M. Ion transport, osmoregulation and acid-base balance. In *The Physiology of Fishes* 177–230 (CRC Press, 2006).
36. Kingsley, D. M. *et al.* New genomic tools for molecular studies of evolutionary change in threespine sticklebacks. *Behaviour* **141**, 1331–1344 (2004).
37. Miyake, A., Mochizuki, S., Yokoi, H., Kohda, M. & Furuichi, K. New ether-à-go-go K<sup>+</sup> channel family members localized in human telencephalon. *J. Biol. Chem.* **274**, 25018–25025 (1999).
38. Miyake, A. *et al.* Disruption of the ether-à-go-go K<sup>+</sup> channel gene BEC1/KCNH3 enhances cognitive function. *J. Neurosci.* **29**, 14637–14645 (2009).
39. Gutman, G. A. *et al.* International Union of Pharmacology. LIII. Nomenclature and molecular relationships of voltage-gated potassium channels. *Pharmacol. Rev.* **57**, 473–508 (2005).
40. Kirkpatrick, M. & Barton, N. Chromosome inversions, local adaptation and speciation. *Genetics* **173**, 419–434 (2006).
41. Hoffmann, A. A. & Rieseberg, L. H. Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? *Annu. Rev. Ecol. Evol. Syst.* **39**, 21–42 (2008).
42. Lowry, D. B. & Willis, J. H. A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biol.* **8**, e1000500 (2010).
43. Joron, M. *et al.* Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* **477**, 203–206 (2011).
44. Feder, J. L., Roethele, J. B., Filchak, K., Niedbalski, J. & Romero-Severson, J. Evidence for inversion polymorphism related to sympatric host race formation in the apple maggot fly, *Rhagoletis pomonella*. *Genetics* **163**, 939–953 (2003).
45. Barrick, J. E. *et al.* Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* **461**, 1243–1247 (2009).
46. Kvitek, D. J. & Sherlock, G. Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape. *PLoS Genet.* **7**, e1002056 (2011).
47. Wittkopp, P. & Haerum, B. K. Regulatory changes underlying expression differences within and between *Drosophila* species. *Nature Genet.* **40**, 346–350 (2008).
48. Reimchen, T. E., Stinson, E. M. & Nelson, J. S. Multivariate differentiation of parapatric and allopatric populations of threespine stickleback in the Sangar River watershed, Queen Charlotte Islands. *Can. J. Zool.* **63**, 2944–2951 (1985).
49. Deagle, B. E. *et al.* Population genomics of parallel phenotypic evolution in stickleback across stream–lake ecological transitions. *Proc. R. Soc. B* **279**, 1277–1286 (2011).
50. McPhail, J. D. Speciation and the evolution of reproductive isolation in the sticklebacks (*Gasterosteus*) of south-western British Columbia. In *The Evolutionary Biology of the Threespine Stickleback* 399–437 (Oxford Univ. Press, 1994).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** Stickleback sequencing at Broad Institute was supported by grants from the National Human Genome Research Institute (NHGRI). R.M.M., J.S., J.G. and D.M.K. were supported by NHGRI CEGS Grant P50-HG002568; Y.F.C. by a Stanford Affymetrix Bio-X Graduate Fellowship; C.T.M. by the Jane Coffins Childs Fund; and B.R.S., T.R.H. and A.A.P. by graduate fellowships from NSF and NDSEG. D.M.K. is an investigator of the Howard Hughes Medical Institute. K.L.-T. is a EURI award recipient funded by ESF. We thank W. Cresko for the BEPA individual used for reference genome sequencing; M. Bell, J. McKinnon, B. Jónsson, S. Mori, C. Peichel, D. Schluter, M. Kalbe, T. Reimchen, D.-P. Højgaard, M. McLaughlin, B. Geyti and B. Blackman for discussions and assistance with specimens used in population surveys; and G. Bejerano for useful discussions and assistance with computational analysis.

**Author Contributions** K.L.-T., F.D.P., E.S.L. and D.M.K. planned and oversaw the project and K.L.-T. and D.M.K. are co-senior authors. E.M. and M.G.G. assembled, J.S., J.G., M.C.D., A.K.K. and R.M.M. anchored, and S.W., E.B. and S.S. annotated the reference genome. C.A. constructed the BEPA BAC library. F.C.J., Y.F.C., D.M.K., K.L.-T., F.D. and M.G.G. designed the whole-genome re-sequencing experiment. F.C.J. and Y.F.C. performed morphometric analyses. S.D.B. and J.J. prepared and coordinated samples. M.G.G., M.P. and M.C.Z. analysed pilot data and performed simulations to evaluate sequencing strategies. M.G.G., P.R., E.M., F.C.J., Y.F.C., J.J. and R.S. analysed polymorphisms. P.R. and M.G.G. developed and carried out the SOM/HMM analysis. F.C.J. and Y.F.C. developed and carried out the CSS and DDC analysis. P.R. analysed gene and non-coding element density, and performed phylogenetic analysis. T.H. analysed GO term enrichments. F.C.J. and Y.F.C. carried out hybrid zone analysis. C.T.M., B.R.S., J.G., J.S., Y.F.C. and F.C.J. analysed chromosome inversions. F.C.J. and D.M.K. performed analysis of coding and regulatory changes. H.Z., A.A.P. and T.H. performed the whole-genome expression study. D.M.K., F.C.J., Y.F.C., P.R., F.D.P. and K.L.-T. wrote the paper with input from other authors.

**Author Information** UCSC Genome browser tracks showing genome-wide analyses are available at <http://sticklebrowser.stanford.edu>. Microarray expression data are deposited at the Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo>) under accession number GSE34783. BAC-end sequences are deposited at <http://www.ncbi.nlm.nih.gov/dbGSS> (accession numbers JS583469 to JS583576). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all readers at [www.nature.com/nature](http://www.nature.com/nature). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to K.L.-T. ([kiersli@broad.mit.edu](mailto:kiersli@broad.mit.edu)) or D.M.K. ([kingsley@stanford.edu](mailto:kingsley@stanford.edu)).

#### Broad Institute Genome Sequencing Platform & Whole Genome Assembly Team

Jen Baldwin<sup>1</sup>, Toby Bloom<sup>1</sup>, David B. Jaffe<sup>1</sup>, Robert Nicol<sup>1</sup> & Jane Wilkinson<sup>1</sup>

<sup>1</sup>Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA.



# Choice-specific sequences in parietal cortex during a virtual-navigation decision task

Christopher D. Harvey<sup>1,3,4†</sup>, Philip Coen<sup>1,4</sup> & David W. Tank<sup>1,2,3,4</sup>

The posterior parietal cortex (PPC) has an important role in many cognitive behaviours; however, the neural circuit dynamics underlying PPC function are not well understood. Here we optically imaged the spatial and temporal activity patterns of neuronal populations in mice performing a PPC-dependent task that combined a perceptual decision and memory-guided navigation in a virtual environment. Individual neurons had transient activation staggered relative to one another in time, forming a sequence of neuronal activation spanning the entire length of a task trial. Distinct sequences of neurons were triggered on trials with opposite behavioural choices and defined divergent, choice-specific trajectories through a state space of neuronal population activity. Cells participating in the different sequences and at distinct time points in the task were anatomically intermixed over microcircuit length scales (<100 micrometres). During working memory decision tasks, the PPC may therefore perform computations through sequence-based circuit dynamics, rather than long-lived stable states, implemented using anatomically intermingled microcircuits.

In real-world tasks, decision-making and working memory often occur in the context of other complex behaviours, including spatial navigation. For example, when driving through a city towards a destination, sensory information defining context and place engages memory and decision circuits to plan turns at upcoming intersections. The PPC is a prime candidate for the neuronal circuitry combining the cognitive processing elements necessary for such tasks. In primates, the PPC is important for perceptual decision-making and categorization<sup>1–3</sup>, movement planning<sup>4</sup> and spatial attention<sup>5</sup>. Studies in rats suggest that the PPC is also important for encoding route progression during navigation<sup>6–9</sup>. Using a virtual-reality system for mice<sup>10</sup> and cellular resolution optical imaging methods<sup>11,12</sup>, we developed a T-maze-based navigation task combining all these cognitive elements and characterized the neuronal circuit dynamics in the PPC, which have not been studied in this combined behavioural context.

Neuronal activity patterns in the PPC have been studied using microelectrode recordings during spatial attention, working memory and perceptual decision tasks. These studies have commonly found cells with sustained firing rate changes spanning entire task periods (cue, delay, response periods)<sup>2,4,5,13</sup>. For example, cells with persistent activity throughout the delay period of memory-guided saccade/reach tasks have frequently been recorded<sup>14,15</sup>. Also, studies have identified neurons with ramps of increasing firing rate spanning the accumulation of evidence period in a motion perception task<sup>1,2</sup>. Neurons with sustained activity during the same task period often have similar activity time courses, suggesting the presence of classes of cells (for example, delay cells) and implying that the task-dependent neuronal dynamics are low-dimensional. The low-dimensional dynamics can be reproduced in recurrent attractor network models, in which each cell's activity is typically an amplitude-scaled version of a prototypical time series<sup>16–21</sup>. In contrast, recent analysis of prefrontal cortex activity has identified heterogeneous neuronal activity time courses, in which a neuron's activity can be thought of as the sum of a few activity modes<sup>22–24</sup>.

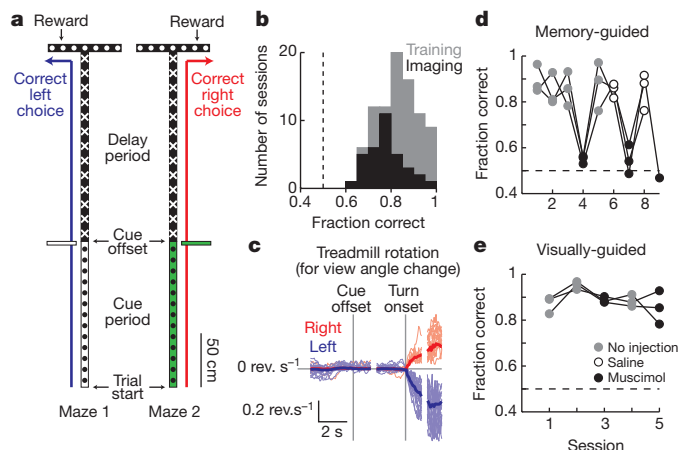
In addition, there is growing evidence for sequences of neuronal activation within local circuits, in which each neuron is active for only a fraction of a task period, including during working memory tasks in the prefrontal cortex<sup>25–28</sup> and the hippocampus<sup>29</sup> and during an object construction task in the PPC<sup>30</sup>. Sequences suggest dynamics that are high-dimensional, without the presence of classes of cells with relatively homogeneous activity time courses.

We explored whether PPC dynamics are best described in terms of cell classes or high-dimensional dynamics. During a navigation-based decision task, the dynamics were high-dimensional: neurons were active in choice-specific sequences in which information moves from one neuronal population to another across time in the task; although the neuronal activity patterns could be divided into cue, delay and response groups, sequences were present within each group. Furthermore, exploiting the ability of cellular resolution optical measurements to provide the relative anatomical location of the recorded cells<sup>12,31,32</sup>, we found that neurons active during behaviourally distinct task periods and on trials with different behavioural choices were spatially intermixed over microcircuit length scales.

## A PPC-dependent navigation-based decision task

Using a virtual-reality system<sup>10</sup>, we trained mice to navigate through a virtual T-maze in a task that incorporated both visual discrimination and a memory-guided response (Methods; Fig. 1a). Visual cues present in the initial section of the maze indicated which direction to turn at the T-intersection to receive a water reward (one set of cues to indicate a right turn and a second set for a left turn; Supplementary Fig. 1c). Between the cue section and the T-intersection, mice ran through a delay maze section that was identical on left and right turn trials; visual information about the reward location was thus present only in the cue section. The task resembled traditional delay tasks<sup>14,15,33</sup> in its cue–delay–response structure, but differed in that continuing sensorimotor activity was present throughout the task, including the delay period.

<sup>1</sup>Princeton Neuroscience Institute, Princeton University, Princeton, New Jersey 08544, USA. <sup>2</sup>Bezos Center for Neural Circuit Dynamics, Princeton University, Princeton, New Jersey 08544, USA. <sup>3</sup>Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey 08544, USA. <sup>4</sup>Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA. <sup>†</sup>Present address: Department of Neurobiology, Harvard Medical School, Boston, Massachusetts 02115, USA.



**Figure 1 | A PPC-dependent decision task in virtual reality.** **a**, Diagram of the two versions of the virtual T-maze that differed only in the cue period and the reward location. Patterns in the diagram reflect the patterns present on the virtual maze walls. **b**, Behavioural performance on individual training (grey) and imaging (black) sessions. **c**, Rotational velocity of the spherical treadmill about the vertical axis for view angle changes on correct right (red) and left (blue) trials, aligned to the cue offset and the turn onset. **d**, Behavioural performance on a memory-guided task from **a** after receiving no injections (grey), saline (open circles) or muscimol (black) bilaterally in the PPC. Connected dots are from individual mice across daily sessions ( $n = 3$  mice). **e**, Same as in **d** except for a visually-guided task (Supplementary Fig. 1b; key in **e** applies to **d** also).

Mice performed the task with high levels of accuracy ( $83 \pm 9\%$  correct,  $P < 0.0001$  versus 50%,  $t$ -test; Fig. 1b). Within and across trials, individual mice ran at highly consistent speeds (Supplementary Fig. 2); however, running speeds varied between mice, and the maze position at which mice began rotating the spherical treadmill to make a right or left turn differed across trials (turn onset; Fig. 1c, Supplementary Fig. 3; Methods). To compare behaviourally similar epochs of the task across trials, individual trials were aligned to the time points when the cue was no longer visible (cue offset), the turn onset, and the end of the trial; these alignment points formed the boundaries of cue (trial start to cue offset,  $5.1 \pm 2.6$  s), delay (cue offset to turn onset,  $4.2 \pm 1.2$  s) and turn task periods (turn onset to trial end,  $3.0 \pm 1.8$  s).

Because the location of the mouse PPC has not been characterized, we first performed retrograde and anterograde labelling experiments to locate it anatomically (Methods, Supplementary Fig. 4). We identified a region consistent with the rat and primate PPC based on the set of areas from where it received axonal projections, the areas to which it sent projections, and its location relative to other cortical regions<sup>8,34</sup> (anterior to visual cortex and posterior to somatosensory cortex). We therefore considered this area to be the mouse PPC.

To test if the PPC was required for the behavioural task, we inactivated it using bilateral injections of the GABA<sub>A</sub> receptor agonist, muscimol. Muscimol reversibly decreased behavioural performance from high levels of accuracy in control sessions to near chance levels, but did not affect the rate of trials performed (fraction correct; no injections  $87 \pm 7\%$ , saline  $85 \pm 5\%$ , muscimol  $54 \pm 5\%$   $P < 0.0001$  versus no injections,  $t$ -test; trials per minute; no injections  $3.2 \pm 0.5$ , muscimol  $3.1 \pm 0.3$ ,  $P > 0.8$ ,  $t$ -test; Fig. 1d). In contrast, PPC inactivation did not significantly affect performance on a visually guided task in which a visual cue was present at the reward site and visible throughout the trial, indicating that the decrease in performance on the memory-guided task was unlikely to be due to a major visual or motor deficit (fraction correct; no injections,  $90 \pm 5\%$ ; muscimol,  $87 \pm 5\%$ ,  $P > 0.2$ ,  $t$ -test; Fig. 1e, Supplementary Fig. 1b).

### Imaging sequences of neuronal activity

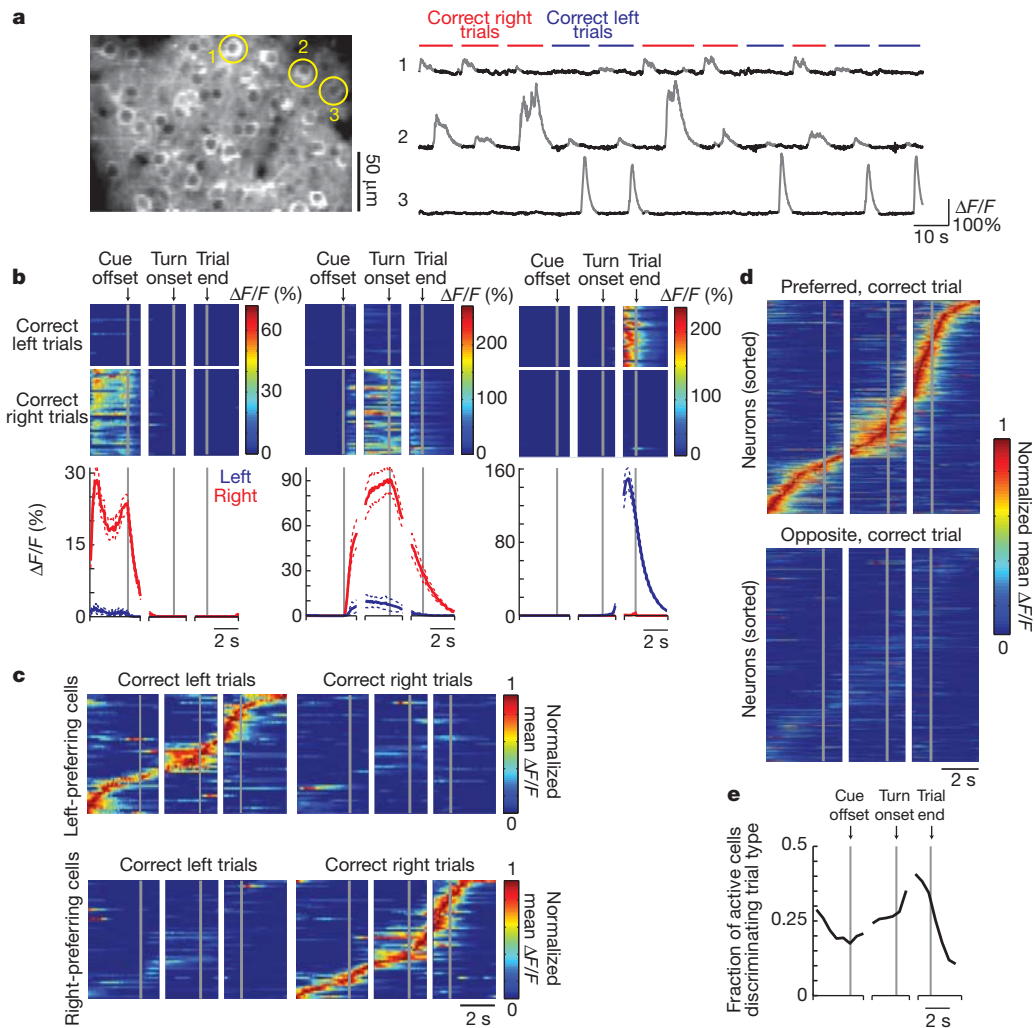
We used two-photon microscopy to image layer 2/3 PPC neurons expressing the genetically encoded calcium indicator GCaMP3, which

increases in fluorescence intensity in response to action potential firing<sup>35</sup> (Methods). On average, we imaged  $\sim 65$  cells simultaneously within an area  $\sim 300 \mu\text{m}$  by  $\sim 150 \mu\text{m}$  (range, 37–94 cells). Nearly all imaged cells showed statistically significant  $\text{Ca}^{2+}$  transients during the behavioural session (96% of cells had  $>0.2$  transients per minute; Methods). Of the cells with high levels of activity ( $>2 \text{ Ca}^{2+}$  transients per minute on average; Supplementary Fig. 5),  $\sim 73\%$  had significant increases in their mean fluorescence intensity traces ( $\Delta F/F$ , averaged across trials) during a specific time in the trial or inter-trial interval (task-modulated cells; Fig. 2, Supplementary Figs 6, 7). These task-modulated cells had  $\text{Ca}^{2+}$  transients for only short time intervals on individual trials ( $11 \pm 8\%$  of time points in trials with a transient) such that only a small fraction of neurons was active simultaneously (Fig. 2b–e, Supplementary Figs 8, 9a); cells with prolonged activity patterns covering a large fraction of the trial were not observed. The majority ( $\sim 71\%$ ) of task-modulated cells had significantly different levels of activity on correct right and left choice trials (choice-specific cells). Similar choice-specific, task-modulated activity patterns were observed in extracellular electrophysiological recordings (Supplementary Figs 10, 11). Cells were also active on error trials, such that neurons active during the cue period tended to be correlated with the cue identity, and neurons active during the turn period in general were correlated with the behavioural response (Supplementary Fig. 12). Only a small fraction of cells had obvious reward-related signals ( $\sim 2\%$  of active cells with  $P < 0.01$ ,  $t$ -test, comparing  $\Delta F/F$  values within  $\sim 0.6$  s after the reward was given on correct trials or missed on error trials).

When the activity patterns of all the choice-specific, task-modulated cells were ordered according to the time profile of their  $\text{Ca}^{2+}$  transients, the active periods across cells were staggered relative to one another in time, forming a sequence of neuronal activation covering the entire trial length (Fig. 2c, d, Supplementary Fig. 7c). Different sequences of neurons were activated on left and right trials (Fig. 2c, d and absence of activity in Fig. 2d lower plot, Supplementary Fig. 14e). Although these plots of sequences combined cells from different experiments and averaged across trials, similar properties were observed when considering only the cells imaged in a single mouse and on individual trials (Fig. 2c, Supplementary Fig. 13). Sequences were also apparent in the  $\sim 29\%$  of task-modulated cells that did not have choice-specific activity; these cells participated in the sequences for both right and left choice trials (Supplementary Fig. 7b). In total,  $\sim 73\%$  of the highly active cells participated in sequences. Sequences similar to those during the task were not observed in shuffled versions of the data set, demonstrating that the sequences were not an artefact created by ordering the data (Supplementary Fig. 14a–d).

Because previous studies of the PPC have categorized cells into classes with cue, delay or response period activity<sup>14,15</sup>, we examined the activity patterns to see if neurons in the sequence were grouped on the basis of behavioural periods. The distribution of activity times of all cells in the population (calculated for each cell as the centre-of-mass (COM) in time of the mean  $\Delta F/F$  during the trial,  $t_{\text{COM}}$ ) had three peaks corresponding to the cue, delay and turn periods, suggesting a possible grouping by behavioural period (Fig. 3a). Consistently, principal component analysis (PCA) of the mean  $\Delta F/F$  traces for all cells revealed three intermixed groups, with each group mostly containing cells preferring the same behavioural period (Fig. 3a, b, Supplementary Fig. 15).

Although the population of neurons could be divided into groups, the temporal activity patterns within each individual period were heterogeneous and formed sequences (Fig. 3c). Cells within their preferred period were active for only a fraction of the period ( $35 \pm 16\%$  of time points in preferred periods with a  $\text{Ca}^{2+}$  transient), with different cells active at different times. Although cells with activity covering a large fraction of the period were occasionally observed (for example, Fig. 2c top panel), these cells were rare (4% of cells with activity lasting for  $>60\%$  of the period; Supplementary Fig. 9b); the distribution of



**Figure 2 | Imaging PPC neuronal activity during the T-maze task.** **a**, Left: example image of GCaMP3-expressing neurons in layer 2/3. Right: example fluorescence intensity traces ( $\Delta F/F$ ; grey portions indicate significant  $\text{Ca}^{2+}$  transients, Methods) for three example cells from the left panel on correct right (red) and left (blue) trials. **b**, Activity patterns during the task for cells 1–3 from **a**. Top: colour-coded  $\Delta F/F$  traces for individual correct left and right choice trials. Each row is a single trial aligned to the cue offset, turn onset and trial end. Bottom: mean  $\Delta F/F$  traces for correct right (red) and left (blue) choice trials. Dashed lines indicate mean  $\pm$  s.e.m. **c**, Normalized mean  $\Delta F/F$  traces for all the

choice-specific, task-modulated cells (one cell per row) imaged in a single mouse and divided by left-prefering ( $n = 51$ ) and right-prefering ( $n = 54$ ) cells. Traces were normalized to the peak of each cell's mean  $\Delta F/F$  trace on preferred trials and sorted by the peak time. Some cells were imaged on different days and in different fields-of-view. **d**, Same as in **c**, except for all mice ( $n = 404$  cells from 6 mice) on preferred and opposite trials. **e**, Fraction of active cells with significantly different activity levels on right and left choice trials as a function of trial time.

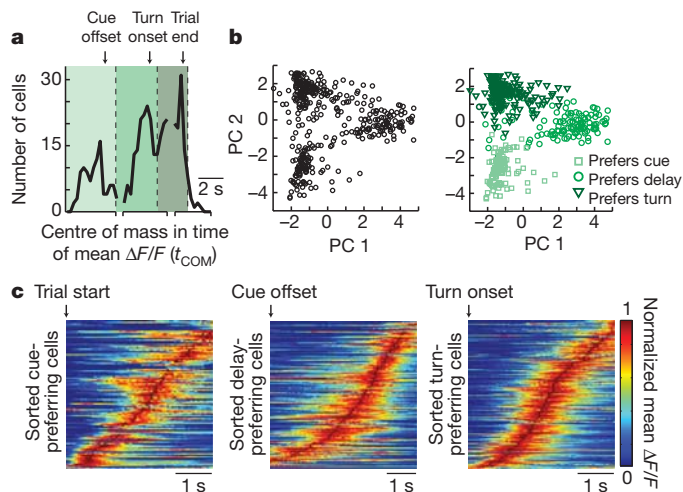
epoch coverage by an individual neuron's activity was similar during the delay and cue or turn periods ( $P > 0.1$ , Kolmogorov-Smirnov test). Furthermore, Pearson's correlations between the non-averaged  $\Delta F/F$  traces for cell pairs with the same trial-type and behavioural period preferences varied widely, with a large fraction of pairs having low correlation coefficients (Supplementary Fig. 16a). The low correlation coefficients could be due to activity at different times in the period, as expected for sequences, or activity at the same time in the trial except on different trials. However, the probability that both cells in these pairs had  $\text{Ca}^{2+}$  transients or that both cells did not have transients on the same trial during their preferred epoch was generally high (Supplementary Fig. 16b), suggesting that the diversity primarily resulted from differences in the activity times of cells within trials. Together these data indicate that classes of cells with homogeneous activity patterns were not present. Rather, choice-specific sequences of neurons were activated in all behavioural periods, with a lower density of cells in the sequence at the borders between periods.

To further examine the sequential neuronal activation on individual trials, we calculated correlations between the activity patterns of cells using the non-averaged  $\Delta F/F$  time series. Cells that were active

at similar time points in the trial on average (measured as the difference in  $t_{\text{COM}}$  values on correct trials,  $\Delta t_{\text{COM}}$ ) had, on correct trials and in their preferred behavioural periods, highly correlated  $\Delta F/F$  traces and peaks in their cross-correlation at a lag approximately equal to  $\Delta t_{\text{COM}}$  (Supplementary Fig. 16d–h); these relationships in the non-averaged  $\Delta F/F$  time series provide further evidence for sequential activity on individual trials. Cells that were active at similar times on correct trials were also highly correlated on error trials and in non-preferred periods (Supplementary Fig. 16c–f). Additionally, cells that were sequentially activated during their preferred behavioural period on correct trials were sequentially active with similar lags during error trials and other time points in the task (Supplementary Fig. 16g–i).

The choice-specific activity could result if mice experienced different visual stimuli and running patterns on right and left trials and if PPC activity was modulated by those differences. To examine this, we first performed a multiple regression analysis to determine the potential effects of the parameters defining the mouse's running trajectory on the fluorescence changes during the delay period (Supplementary Table 1). These parameters could not explain the choice-specific





**Figure 3 | Neuronal activity in individual behavioural periods.** **a**, Histogram of the times of the centre-of-mass of the mean  $\Delta F/F$  trace ( $t_{COM}$ ) for choice-specific, task-modulated cells. Cells were separated into three groups (cue-, delay- and turn-prefering cells; varying shades of green) based on peaks in the distribution. **b**, PCA on the normalized mean  $\Delta F/F$  traces for all the choice-specific, task-modulated cells. Left, scores for each cell plotted for the first two principal components (PC 1, 2). Right, cells categorized on the basis of groupings from **a**.  $n = 404$  cells. **c**, Sorted normalized mean  $\Delta F/F$  traces for cue-prefering ( $n = 101$ ), delay-prefering ( $n = 133$ ) and turn-prefering ( $n = 170$ ) cells, aligned to the trial start, cue offset and turn onset, respectively, on the preferred trial-type.

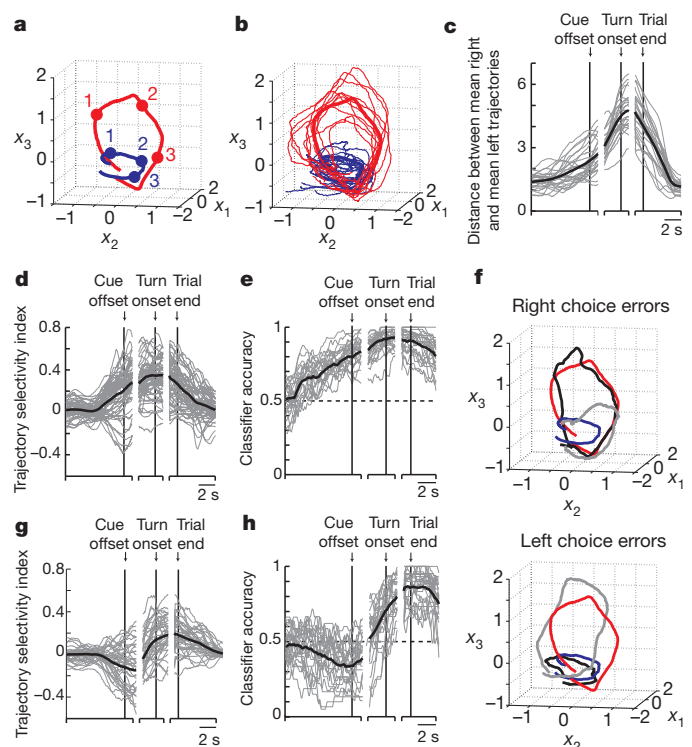
activity patterns, suggesting that any differences in running trajectories between right and left trials did not trigger the activity we observed. In addition, we performed two sets of experiments to further examine whether the maze visual stimuli alone or the mouse's running patterns triggered PPC activity. In the first experiments, movies of simulated left and right turn runs through the T-maze, which closely approximated real runs, were played to mice that passively viewed the visual scenes (open-loop experiments; Supplementary Fig. 17a, b). In the second experiments, mice were trained on a virtual linear track to perform a simple running back and forth task (run to one end for a reward, turn around, run to the other end for the next reward). The linear track had several visual patterns on the walls, all of which were the same as patterns in portions of the T-maze (Supplementary Fig. 18a, b). The task required similar running and turning behaviours to the T-maze, except that turns were not memory-guided based on visual cues and a delay period, and there was not a two-alternative forced-choice structure. The overall levels of activity during the simulated T-maze runs or during the linear track task were much lower than when mice actively performed the T-maze task (Supplementary Figs 17c, d, 18c, d). Also, only a small fraction of neurons had significant increases in activity at specific locations in the maze, either during the simulated T-maze runs or in the linear track, suggesting that cells were not activated robustly by location-specific visual scenes or running patterns (simulated runs, 1.3% of neurons; linear track, 5.8%; T-maze, 32.3%; Supplementary Figs 17e–k, 18e–k). Together these results suggest that PPC neurons in the T-maze were not activated only by the visual information or by the running patterns of the mouse.

### Choice-specific neuronal circuit trajectories

The heterogeneous and sequential neuronal activity patterns during the T-maze task indicated that we should consider the dynamics of the population rather than classes of cells. We therefore analysed the dynamics as a trajectory through a state space of neuronal population activity (neuronal circuit trajectory)<sup>36–38</sup>. At each time point, the activity state of the circuit containing  $n$  simultaneously imaged neurons was defined as a point in an  $n$ -dimensional space, with each dimension representing the activity ( $\Delta F/F$  values) of a single neuron. Different

trajectories (visualized using factor analysis for dimensionality reduction) were traversed for trials with different behavioural choices (Fig. 4a, b, Supplementary Fig. 19a). The trajectories for correct right and left choice trials began the trial at similar positions, gradually diverged to reach a peak separation near the time of the behavioural choice, and converged to the starting point in the inter-trial interval (Fig. 4a–d). To quantify the trajectory divergence, we used a classifier based on the distance from an individual trial trajectory to the mean right and left choice trajectories at single time points. It was possible, from the activity of a small population of neurons located in close anatomical proximity ( $\sim 65$  neurons separated by  $< 250 \mu\text{m}$ ), to predict the mouse's choice on single correct trials at better than chance levels during the cue, delay and turn periods (Methods; Fig. 4e, Supplementary Fig. 19b, c). The activity in the PPC can therefore be considered as divergent, choice-specific trajectories through a state space of neuronal population activity.

Trajectories were highly variable on error trials. Some trajectories began close to the correct choice trajectory during the cue period and

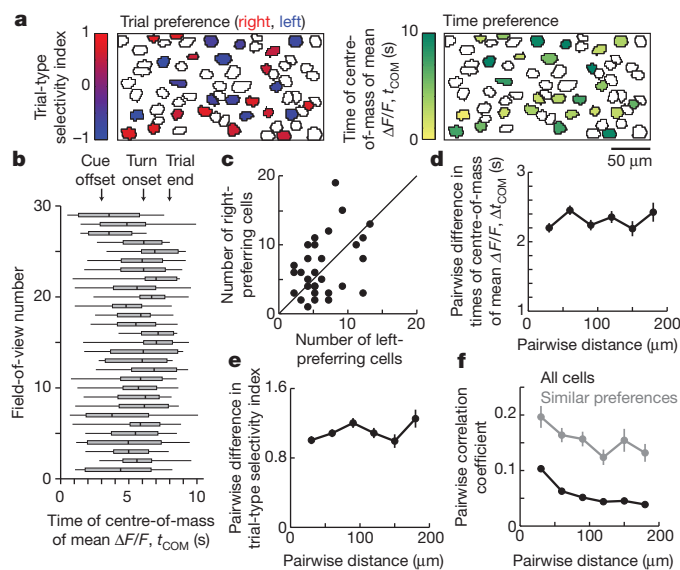


**Figure 4 | Neuronal circuit trajectories on correct and error trials.** **a**, Time course of mean choice-specific trajectories on correct right (red) and left (blue) choice trials from one session (plotted for the first three common factors). Points labelled 1, 2 and 3 correspond to the mean times of the cue offset, turn onset and trial end, respectively. **b**, Example individual (thin lines) and mean (thick lines) trajectories for correct trials from the session in **a**. **c**, Euclidean distance between the mean trajectories on correct right and left choice trials ( $n = 29$  individual sessions, grey). The black line indicates the mean. **d**, Trajectory selectivity index for individual correct trials from a single session, defined on the basis of distances as  $(d_{\text{to mean traj, opposite choice}} - d_{\text{to mean traj, same choice}}) / (d_{\text{to mean traj, same choice}} + d_{\text{to mean traj, opposite choice}})$ . Values close to 1 and  $-1$  indicate that the individual trial trajectory was near the mean trajectory of the same and the opposite behavioural choice, respectively. Mean trajectories were defined using correct trials only. Grey and black lines indicate individual trials and the mean, respectively. **e**, Classification accuracy for determining the behavioural choice of the mouse at different time points in the task during individual sessions (grey,  $n = 29$ ; black, mean). The classifier was based on a distance-dependent classification scheme (see Methods) using correct trials. **f**, Example individual trial trajectories (grey and black) on right choice and left choice error trials, plotted with the mean trajectories for correct right (red) and left (blue) choice trials. **g**, Same as for **d**, except on error trials. **h**, Same as for **e**, except on error trials.

transitioned towards the error choice trajectory later in the trial; such transitions occurred at a wide range of points in the trial (grey traces in Fig. 4f, Supplementary Fig. 19f). Other trajectories were similar to the error choice trajectory throughout the trial (Fig. 4f, black traces). On average, the error trial trajectories were more similar to the correct choice trajectory during the cue period and closer to the error choice trajectory during the turn period (cue offset; trajectory selectivity index =  $-0.13 \pm 0.22$ ,  $P < 0.001$  versus 0, *t*-test: trial end; trajectory selectivity index =  $0.18 \pm 0.15$ ,  $P < 0.001$  versus 0; Fig. 4g; consistent results based on classification, Fig. 4h). Therefore, individual trajectories transitioned between the mean correct right and left choice trajectories at many time points during trials, but most frequently switched during the delay period.

### Anatomical micro-organization of PPC dynamics

To determine how the cells implementing the activity dynamics were anatomically organized, we first examined the neuronal activity patterns to see if cells with different response preferences, such as activity peaks at different times in the trial or different preferences for right and left choice trials, were present in the same regions of the PPC or separated into different areas. Each  $\sim 250 \mu\text{m}$  by  $\sim 125 \mu\text{m}$  area (that is, field-of-view with simultaneously imaged cells) contained both right and left choice-prefering cells of approximately equal numbers and cells with activity peaks at a wide range of times in the trial (Fig. 5a–c, Supplementary Fig. 20). Next, within each imaged area we compared the activities of pairs of neurons as a function of the distance between the neurons' cell bodies. The difference in the trial-type selectivity for cells in a pair did not depend significantly on the distance between cells, indicating that left and right choice-prefering cells were intermixed ( $\rho = 0.04$ , Spearman's correlation,  $P > 0.05$ ; Fig. 5e). Similarly,  $\Delta t_{\text{COM}}$



**Figure 5 | Anatomical micro-organization in the PPC** **a**, Example field-of-view with cells outlined and choice-specific, task-modulated cells coloured. Left, cells' trial-type selectivity, defined as  $(\Delta F/F_{\text{right trials}} - \Delta F/F_{\text{left trials}})/(\Delta F/F_{\text{right trials}} + \Delta F/F_{\text{left trials}})$ . Values close to 1 (red) and  $-1$  (blue) indicate right and left choice preferences, respectively. Right, cells' time of the centre-of-mass of the mean  $\Delta F/F$  trace ( $t_{\text{COM}}$ ).  $t_{\text{COM}} = 0$  corresponds to  $\sim 3$  s before the cue offset. **b**, Box plots of  $t_{\text{COM}}$  values for task-modulated cells in individual fields-of-view (box edges, first and third quartiles; vertical line in the box, median; whiskers, range) **c**, Number of right and left choice-prefering cells in each field-of-view ( $n = 29$ ; unity line is shown). **d**, Difference in  $t_{\text{COM}}$  as a function of the distance between cells. **e**, Difference in the trial-type selectivity index as a function of the distance between cells. **f**, Pearson's correlation between non-averaged  $\Delta F/F$  traces (all time points) for all pairs of active cells (black) and cell pairs with the same trial-type and behavioural period preference (grey) as a function of the cell-cell distance.

for a cell pair did not differ with the distance between the cells, indicating that cells active during different periods of the task were intermingled ( $\rho = -0.01$ ,  $P > 0.6$ ; Fig. 5d). However, cells that were separated by less anatomical distance had  $\Delta F/F$  traces that were significantly more correlated than cell pairs further apart, but the relationship between the correlation coefficient and distance was weak (all pairs of active cells;  $\rho = -0.16$ ,  $P < 0.001$ ; pairs with the same trial-type and behavioural period preference;  $\rho = -0.22$ ,  $P < 0.001$ ; Fig. 5f). This weak relationship could be due to overlapping fluorescence changes, such as from dendritic signals, or could reflect an actual, weak spatial organization. Regardless, cells with highly different activity patterns were intermixed over short length scales, and an anatomical separation of the response properties we measured was not present.

### Discussion

The choice-specific sequences of PPC neuronal activation we report here add to the growing list of studies that have identified cortical sequences of activity states during working memory tasks<sup>27–30</sup>. Furthermore, because we demonstrated that PPC activity was necessary only for the memory-guided task, the sequences of activation were probably important at least for the memory aspect of the task. Sequence-based dynamics may therefore be a common framework for circuit function during memory and decision tasks, including during navigation behaviours. Such dynamics could potentially be implemented using feedforward architectures<sup>39,40</sup> or liquid state machines<sup>41,42</sup> related to those that have been proposed for working memory.

Our results also offer a way to unite previous work on neural coding in the PPC. Navigation, memory and choice information may be combined in the sequences such that the identity of the active sequence reflects choice-related information for working memory and movement planning<sup>4,5</sup>, and that the currently active cell within the sequence reflects spatial or temporal progression through the task<sup>7,43</sup>, which were highly correlated in our task (Supplementary Fig. 2). It seems unlikely that PPC neurons only provided location information in a context-dependent manner, like hippocampal place cells<sup>29,44</sup>, because unlike place cells PPC neurons did not encode spatial location during a linear track task (Supplementary Fig. 18) or other tasks<sup>6,7</sup>.

A possible explanation for heterogeneity and sequences in our experiments versus stereotypy and low-dimensional dynamics, which have been emphasized in previous recordings and models of PPC activity<sup>1,13–18,20,21</sup>, is that the PPC adopts different dynamics depending on the demands of the behavioural task. Because traditional delayed saccade tasks, for example, have one spatiotemporal component during the delay period (fixation before making a response), the PPC may adopt sustained activity patterns. In contrast, during tasks that involve many spatial and temporal components, as are common in natural behaviours and during navigation, the PPC may utilize sequences of activation. Alternatively, sequences of activity may be present in the primate PPC during traditional tasks but have yet to be identified, consistent with emerging evidence for heterogeneous temporal response properties<sup>22–24,30</sup>. In addition, different regions or layers of the PPC may have differing activity dynamics<sup>45</sup>, or the dynamics of rodent and primate PPC circuits may differ.

Because cells that were active at distinct time points in the task and that participated in different choice-specific sequences were spatially intermingled, our results indicate that functionally distinct sub-networks are anatomically interlaced in the PPC. This extends previous work in sensory cortex, motor cortex and the hippocampus showing spatial intermixing of heterogeneous response properties in cells encoding qualitatively similar types of information (for example, orientation selectivity in visual cortex) or in cells with activity during similar task epochs<sup>12,31,32</sup>. Our findings differ from the predictions of models that have emphasized the spatial clustering of similar response patterns, as in functional columns, and that propose connectivity defined by axonal–dendritic overlap without fine-scale specificity<sup>46,47</sup>. Rather, our results support a model in which microcircuits are formed



by highly specific synaptic connectivity and are composed of neuronal motifs, such as those identified in brain slice recordings in sensory cortices<sup>48,49</sup> and amongst visual cortical neurons with the same orientation selectivity<sup>50</sup>. Our results showing that cell pairs that were sequentially active during their preferred periods on correct trials had similar activity relationships even during error trials and the inter-trial interval suggest the presence of such motifs (Supplementary Fig. 16g-i).

The behavioural task used here did not isolate the decision-making process. We note however that activity trajectories occasionally switched during a trial between the prototypical correct left and right choice trajectories, including frequently on error trials, suggesting that the mouse's decision was not necessarily irreversibly reached immediately after a trial's start (trajectory selectivity switch during the delay period or last 1 s of the cue period; 63% of error trials, 20% of correct trials; Fig. 4d, g, Supplementary Fig. 19e, f). Sequences may therefore play a role in some aspect of decision-related processes, but further experiments will be necessary to assess this possibility.

Our results motivate consideration of a conceptual framework for decision-making and working memory in which sensory information used for the decision activates a neuronal sequence of activity. The sequence begins in a choice-independent state, which could be mediated by neurons that are not choice-specific (Supplementary Fig. 7b), and then moves towards a choice-specific trajectory and away from other trajectories in a manner dependent on the incoming information. A decision is proposed to be reached when the sequence of activity intersects a choice-specific trajectory; different decisions involve intersections with different trajectories. Upon reaching a decision, a working memory can be maintained by continuing along that choice-specific trajectory. Changing decisions would occur through transitions between trajectories, but as time progresses in the task, the state space distance between trajectories increases, in effect creating a larger barrier to change. In this view, decision-making and working memory utilize an ordered progression through a sequence in which information moves from one population of neurons to another over time. The framework we propose is an extension of a point-of-view first considered in describing the dynamics underlying behavioural choices in the leech nervous system<sup>37</sup>. It has some similarities with (and some differences from) drift/diffusion-to-bound, race, and recurrent network models of decision-making implemented as neuronal integrator winner-take-all circuits<sup>2,16</sup>. These models are similar to the trajectory-based view in that different decisions correspond to a divergence in state space surrounding a separatrix. However, these models differ from the sequence framework in that decisions are reached when the activity approaches a choice-specific fixed point with working memory maintained as stable activity at that point. Thus, although these circuits can demonstrate divergent trajectories to reach the fixed points associated with different choices, the trajectories are defined by relatively homogeneous changes in the activity of the population, and the same set of neurons participate in all stages of the decision-making and working memory process.

## METHODS SUMMARY

Using a previously described virtual-reality system<sup>10</sup>, male C57/BL6 mice were trained using operant conditioning to navigate through a virtual T-maze to receive water rewards. Translation and rotation in the virtual environment were controlled by the mouse's running on a spherical treadmill. Training was performed using shaping implemented as a set of six mazes of increasing task difficulty. Retrograde tracing was performed using fluorescent beads, and anterograde tracing was performed following injections of adeno-associated virus (AAV) containing GFP or GCaMP3. Muscimol injections for PPC inactivation were made bilaterally ~350  $\mu\text{m}$  beneath the dura (50 nl, 1 ng nl<sup>-1</sup>). Imaging was performed using a custom two-photon microscope incorporated with the virtual-reality system, as described previously<sup>12</sup>. Imaging occurred at 2–6 weeks after injection of AAV2/1-synapsin-1-GCaMP3 virus<sup>35</sup>. A complete description of the experimental methods and data analysis is available in the Supplementary Information.

Received 5 December 2011; accepted 2 February 2012.

Published online 14 March 2012.

- Shadlen, M. N. & Newsome, W. T. Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *J. Neurophysiol.* **86**, 1916–1936 (2001).
- Gold, J. I. & Shadlen, M. N. The neural basis of decision making. *Annu. Rev. Neurosci.* **30**, 535–574 (2007).
- Freedman, D. J. & Assad, J. A. A proposed common neural mechanism for categorization and perceptual decisions. *Nature Neurosci.* **14**, 143–146 (2011).
- Andersen, R. A. & Cui, H. Intention, action planning, and decision making in parietal-frontal circuits. *Neuron* **63**, 568–583 (2009).
- Bisley, J. W. & Goldberg, M. E. Attention, intention, and priority in the parietal lobe. *Annu. Rev. Neurosci.* **33**, 1–21 (2010).
- McNaughton, B. L. *et al.* Cortical representation of motion during unrestrained spatial navigation in the rat. *Cereb. Cortex* **4**, 27–39 (1994).
- Nitz, D. A. Tracking route progression in the posterior parietal cortex. *Neuron* **49**, 747–756 (2006).
- Whitlock, J. R., Sutherland, R. J., Witter, M. P., Moser, M. B. & Moser, E. I. Navigating from hippocampus to parietal cortex. *Proc. Natl Acad. Sci. USA* **105**, 14755–14762 (2008).
- Calton, J. L. & Taube, J. S. Where am I and how will I get there from here? A role for posterior parietal cortex in the integration of spatial information and route planning. *Neurobiol. Learn. Mem.* **91**, 186–196 (2009).
- Harvey, C. D., Collman, F., Dombeck, D. A. & Tank, D. W. Intracellular dynamics of hippocampal place cells during virtual navigation. *Nature* **461**, 941–946 (2009).
- Dombeck, D. A., Khabbazi, A. N., Collman, F., Adelman, T. L. & Tank, D. W. Imaging large-scale neural activity with cellular resolution in awake, mobile mice. *Neuron* **56**, 43–57 (2007).
- Dombeck, D. A., Harvey, C. D., Tian, L., Looger, L. L. & Tank, D. W. Functional imaging of hippocampal place cells at cellular resolution during virtual navigation. *Nature Neurosci.* **13**, 1433–1440 (2010).
- Curtis, C. E. & Lee, D. Beyond working memory: the role of persistent activity in decision making. *Trends Cogn. Sci.* **14**, 216–222 (2010).
- Barash, S., Bracewell, R. M., Fogassi, L., Gnadt, J. W. & Andersen, R. A. Saccade-related activity in the lateral intraparietal area. I. Temporal properties; comparison with area 7a. *J. Neurophysiol.* **66**, 1095–1108 (1991).
- Chafee, M. V. & Goldman-Rakic, P. S. Matching patterns of activity in primate prefrontal area 8a and parietal area 7ip neurons during a spatial working memory task. *J. Neurophysiol.* **79**, 2919–2940 (1998).
- Wang, X. J. Decision making in recurrent neuronal circuits. *Neuron* **60**, 215–234 (2008).
- Wong, K. F. & Wang, X. J. A recurrent network mechanism of time integration in perceptual decisions. *J. Neurosci.* **26**, 1314–1328 (2006).
- Mazurek, M. E., Roitman, J. D., Ditterich, J. & Shadlen, M. N. A role for neural integrators in perceptual decision making. *Cereb. Cortex* **13**, 1257–1269 (2003).
- Ganguli, S. *et al.* One-dimensional dynamics of attention and decision making in LIP. *Neuron* **58**, 15–25 (2008).
- Miller, P., Brody, C. D., Romo, R. & Wang, X. J. A recurrent network model of somatosensory parametric working memory in the prefrontal cortex. *Cereb. Cortex* **13**, 1208–1218 (2003).
- Machens, C. K., Romo, R. & Brody, C. D. Flexible control of mutual inhibition: a neural model of two-interval discrimination. *Science* **307**, 1121–1124 (2005).
- Machens, C. K., Romo, R. & Brody, C. D. Functional, but not anatomical, separation of “what” and “when” in prefrontal cortex. *J. Neurosci.* **30**, 350–360 (2010).
- Jun, J. K. *et al.* Heterogeneous population coding of a short-term memory and decision task. *J. Neurosci.* **30**, 916–929 (2010).
- Singh, R. & Eliasmith, C. Higher-dimensional neurons explain the tuning and dynamics of working memory cells. *J. Neurosci.* **26**, 3667–3678 (2006).
- Batuev, A. S. Two neuronal systems involved in short-term spatial memory in monkeys. *Acta Neurobiol. Exp. (Warsz.)* **54**, 335–344 (1994).
- Seidemann, E., Meilijson, I., Abeles, M., Bergman, H. & Vaadia, E. Simultaneously recorded single units in the frontal cortex go through sequences of discrete and stable states in monkeys performing a delayed localization task. *J. Neurosci.* **16**, 752–768 (1996).
- Baeg, E. H. *et al.* Dynamics of population code for working memory in the prefrontal cortex. *Neuron* **40**, 177–188 (2003).
- Fujisawa, S., Amarasingham, A., Harrison, M. T. & Buzsaki, G. Behavior-dependent short-term assembly dynamics in the medial prefrontal cortex. *Nature Neurosci.* **11**, 823–833 (2008).
- Pastalkova, E., Itskov, V., Amarasingham, A. & Buzsaki, G. Internally generated cell assembly sequences in the rat hippocampus. *Science* **321**, 1322–1327 (2008).
- Crowe, D. A., Averbeck, B. B. & Chafee, M. V. Rapid sequences of population activity patterns dynamically encode task-critical spatial information in parietal cortex. *J. Neurosci.* **30**, 11640–11653 (2010).
- Ohki, K., Chung, S., Ch'ng, Y. H., Kara, P. & Reid, R. C. Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. *Nature* **433**, 597–603 (2005).
- Komiyama, T. *et al.* Learning-related fine-scale specificity imaged in motor cortex circuits of behaving mice. *Nature* **464**, 1182–1186 (2010).
- Erich, J., Bialek, M. & Brody, C. D. A cortical substrate for memory-guided orienting in the rat. *Neuron* **72**, 330–343 (2011).



34. Corwin, J. V. & Reep, R. L. Rodent posterior parietal cortex as a component of a cortical network mediating directed spatial attention. *Psychobiology* **26**, 87–102 (1998).
35. Tian, L. *et al.* Imaging neural activity in worms, flies and mice with improved GCaMP calcium indicators. *Nature Methods* **6**, 875–881 (2009).
36. Mazor, O. & Laurent, G. Transient dynamics versus fixed points in odor representations by locust antennal lobe projection neurons. *Neuron* **48**, 661–673 (2005).
37. Briggman, K. L., Abarbanel, H. D. & Kristan, W. B. Jr. Optical imaging of neuronal populations during decision-making. *Science* **307**, 896–901 (2005).
38. Churchland, M. M., Yu, B. M., Sahani, M. & Shenoy, K. V. Techniques for extracting single-trial activity patterns from large-scale neural recordings. *Curr. Opin. Neurobiol.* **17**, 609–618 (2007).
39. Goldman, M. S. Memory without feedback in a neural network. *Neuron* **61**, 621–634 (2009).
40. Ganguli, S., Huh, D. & Sompolinsky, H. Memory traces in dynamical systems. *Proc. Natl Acad. Sci. USA* **105**, 18970–18975 (2008).
41. Maass, W., Joshi, P. & Sontag, E. D. Computational aspects of feedback in neural circuits. *PLOS Comput. Biol.* **3**, e165 (2007).
42. Sussillo, D. & Abbott, L. F. Generating coherent patterns of activity from chaotic neural networks. *Neuron* **63**, 544–557 (2009).
43. Leon, M. I. & Shadlen, M. N. Representation of time by neurons in the posterior parietal cortex of the macaque. *Neuron* **38**, 317–327 (2003).
44. Wood, E. R., Dudchenko, P. A., Robitsek, R. J. & Eichenbaum, H. Hippocampal neurons encode information about different types of memory episodes occurring in the same location. *Neuron* **27**, 623–633 (2000).
45. Burke, S. N. *et al.* Differential encoding of behavior and spatial context in deep and superficial layers of the neocortex. *Neuron* **45**, 667–674 (2005).
46. Braitenberg, V. B. & Schuz, A. *Anatomy of the Cortex: Statistics and Geometry* (Springer, 1991).
47. Binzegger, T., Douglas, R. J. & Martin, K. A. A quantitative map of the circuit of cat primary visual cortex. *J. Neurosci.* **24**, 8441–8453 (2004).
48. Yoshimura, Y., Dantzker, J. L. & Callaway, E. M. Excitatory cortical neurons form fine-scale functional networks. *Nature* **433**, 868–873 (2005).
49. Song, S., Sjöström, P. J., Reigl, M., Nelson, S. & Chklovskii, D. B. Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biol.* **3**, e68 (2005).
50. Ko, H. *et al.* Functional specificity of local synaptic connections in neocortical networks. *Nature* **473**, 87–91 (2011).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank D. Dombeck for assistance with imaging methods and analysis; C. Domnisoru, M. de Bettencourt, C. Brody and A. Miri for discussions; and M. Goldman, J. Hopfield, D. Aronov, B. Scott and T. Hanks for comments on the manuscript. This work was supported by the NIH (R01-MH083686; RC1-NS068148), a fellowship from the Helen Hay Whitney Foundation (C.D.H.), and a Burroughs Wellcome Fund Career Award at the Scientific Interface (C.D.H.).

**Author Contributions** C.D.H. performed experiments with assistance from P.C. on the retrograde tracing experiments; D.W.T. implemented the imaging instrumentation; C.D.H. analysed the data with strategy and methods contributions from D.W.T.; C.D.H. and D.W.T. wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to C.D.H. ([christopher\\_harvey@hms.harvard.edu](mailto:christopher_harvey@hms.harvard.edu)) or D.W.T. ([dwtank@princeton.edu](mailto:dwtank@princeton.edu)).

# DNA damage defines sites of recurrent chromosomal translocations in B lymphocytes

Ofir Hakim<sup>1\*</sup>, Wolfgang Resch<sup>2\*</sup>, Arito Yamane<sup>2\*</sup>, Isaac Klein<sup>3</sup>, Kyong-Rim Kieffer-Kwon<sup>2</sup>, Mila Jankovic<sup>3</sup>, Thiago Oliveira<sup>3,4</sup>, Anne Bothmer<sup>3</sup>, Ty C. Voss<sup>1</sup>, Camilo Ansarah-Sobrinho<sup>2</sup>, Ewy Mathe<sup>5</sup>, Genqing Liang<sup>2</sup>, Jesse Cobell<sup>2</sup>, Hirotaka Nakahashi<sup>2</sup>, Davide F. Robbiani<sup>3</sup>, Andre Nussenzweig<sup>6</sup>, Gordon L. Hager<sup>1</sup>, Michel C. Nussenzweig<sup>3,7\*</sup> & Rafael Casellas<sup>2,8\*</sup>

**Recurrent chromosomal translocations underlie both haematopoietic and solid tumours. Their origin has been ascribed to selection of random rearrangements, targeted DNA damage, or frequent nuclear interactions between translocation partners; however, the relative contribution of each of these elements has not been measured directly or on a large scale. Here we examine the role of nuclear architecture and frequency of DNA damage in the genesis of chromosomal translocations by measuring these parameters simultaneously in cultured mouse B lymphocytes. In the absence of recurrent DNA damage, translocations between *Igh* or *Myc* and all other genes are directly related to their contact frequency. Conversely, translocations associated with recurrent site-directed DNA damage are proportional to the rate of DNA break formation, as measured by replication protein A accumulation at the site of damage. Thus, non-targeted rearrangements reflect nuclear organization whereas DNA break formation governs the location and frequency of recurrent translocations, including those driving B-cell malignancies.**

Most cancers bear cytogenetic abnormalities including chromosomal translocations and rearrangements<sup>1</sup>. Although translocations and rearrangements are central to the development of cancer, their origins are poorly understood. One possibility is that they arise from rare and random events that are selected in tumour precursors because they provide a growth advantage. However, increasing evidence indicates that mechanistic factors other than simple selection may have a role in their genesis. In B lymphocytes, V(D)J recombination, class switch recombination (CSR) and somatic hypermutation (SHM) produce obligate single- and double-strand DNA break intermediates that can become substrates for translocations<sup>2,3</sup>. Consistent with this idea, genetic ablation of the enzymes that create DNA lesions during V(D)J recombination (RAGs) or CSR and SHM (AID; also called AICDA) has a profound protective effect on B-cell transformation<sup>2,4</sup>.

A second mechanism that may also influence the incidence of chromosomal translocations is nuclear architecture. Two decades of imaging and recent molecular approaches have established that the spatial organization of the genome is not random, but compartmentalized into chromosome territories as well as transcriptionally active and silent subnuclear environments<sup>5–8</sup>. These compartments are believed to influence the frequency with which genes from different chromosomes can interact and recombine. Furthermore, there is a strong association between transcriptional activity and translocation<sup>9</sup>.

Using new methods that capture rearrangements genome-wide, thousands of translocations were recently isolated in primary B cells in the absence of growth selection<sup>9,10</sup>. The studies confirmed the notion that the formation of chromosomal translocations is influenced by spatial conformation, targeted DNA damage and open chromatin.

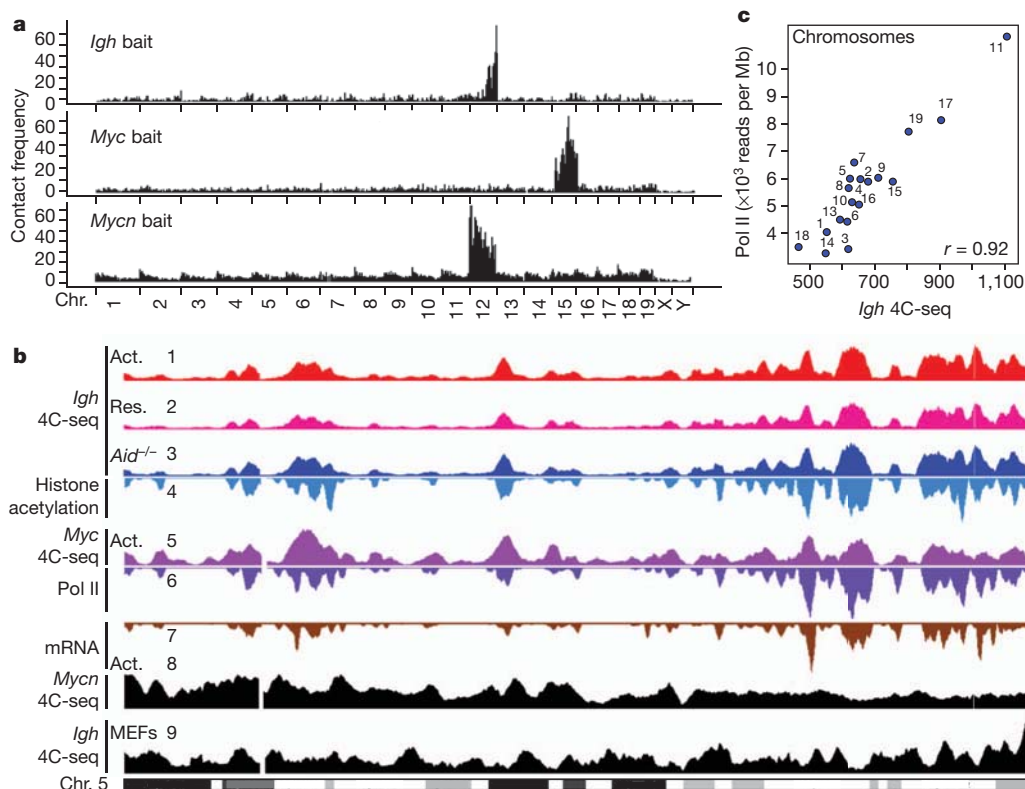
Consistent with the distribution of mammalian chromosomes in discrete nuclear territories, most rearrangements occurred intra-chromosomally<sup>9,10</sup>. Moreover, rearrangements in *trans* were biased towards transcriptionally active genes, and particularly those targeted by AID<sup>9,10</sup>. What the studies did not resolve, however, was to what extent recurrent DNA damage, chromatin accessibility, or spatial genome organization influence the location and frequency of cancer-inducing translocations. Here we make use of deep-sequencing techniques to establish the relationship between genome-wide spatial interactions, DNA damage and translocations in activated B cells.

## A map of *Igh* and *Myc* long-range nuclear associations

To identify genomic regions that are in close spatial proximity to *Igh*, *Myc* and *Mycn* (also called *N-myc*) loci, we performed chromosome conformation capture experiments<sup>11</sup> followed by deep-sequencing (4C-seq). We used *Igh* and *Myc* as baits because they are actively transcribed and targeted by AID<sup>12</sup>. As controls, we analysed *Mycn*, which is transcriptionally silent in peripheral B cells and does not recruit AID<sup>13</sup>, and *Igh* in mouse embryonic fibroblasts (MEFs), where immunoglobulin genes are not expressed. Because of the large size of *Igh*, we used two 4C-seq baits specific for 5'E $\mu$  and 3'E $\alpha$  enhancers (Supplementary Fig. 1a). Two independent 4C libraries (HindIII and BglII) were constructed for each condition (see Methods). In all experiments, most of the 4C sequence reads (76% on average) originated from the *cis* chromosome (Fig. 1a, Supplementary Table 1 and Supplementary Fig. 2a), an observation consistent with the finding that loci on the same chromosome preferentially interact in *cis* within a chromosome territory<sup>5,6,9</sup>.

<sup>1</sup>Laboratory of Receptor Biology and Gene Expression, NCI, National Institutes of Health, Bethesda, Maryland 20892, USA. <sup>2</sup>Genomics & Immunity, NIAMS, NCI, National Institutes of Health, Bethesda, Maryland 20892, USA. <sup>3</sup>Laboratory of Molecular Immunology, The Rockefeller University, New York, New York 10065, USA. <sup>4</sup>Medical School of Ribeirao Preto/USP, Departamento de Genetica, 8 National Institute of Science and Technology for Stem Cells and Cell Therapy and Center for Cell-based Therapy, Ribeirao Preto, SP 14051-140, Brazil. <sup>5</sup>Biodata Mining and Discovery, NIAMS, National Institutes of Health, Bethesda, Maryland 20892, USA. <sup>6</sup>Laboratory of Genome Integrity, NCI, National Institutes of Health, Bethesda, Maryland 20892, USA. <sup>7</sup>Howard Hughes Medical Institute, The Rockefeller University, New York, New York 10065, USA. <sup>8</sup>Center of Cancer Research, NCI, National Institutes of Health, Bethesda, Maryland 20892, USA.

\*These authors contributed equally to this work.



**Figure 1 | Characterization of the *Igh*, *Myc* and *Mycn* interactomes in B lymphocytes.** **a**, Genome-wide interaction profile of *Igh* 3'Ex, *Myc* and *Mycn* in activated B cells. Plots show the percentage of HindIII fragments carrying 4C-seq reads. **b**, Contact frequency of *Igh* with chromosome 5 in activated B cells (lane 1), anti-HEL homozygous activated B cells (lane 2), or resting B cells

(lane 3). Histone acetylation (lane 4), RNA Pol II (lane 6) and mRNA (lane 7) density is also shown. Lanes 5 and 8 show *Myc* and *Mycn* contacts, respectively. Lane 9 represents *Igh* contacts in MEFs. **c**, Comparison of Pol II and *Igh* 4C-seq data per chromosome normalized as reads per mappable megabase.

To explore contact frequencies in *trans*, the mouse genome was partitioned into 200-kilobase (kb) non-overlapping windows and the number of 4C-seq-positive fragments was calculated for each window (see Methods). In activated B cells the contact profile of *Igh* was nonrandom, following a peaks-and-valley pattern similar to that reported for transcriptionally active loci in other cell types<sup>14</sup> (Fig. 1b, lane 1). This pattern was comparable for  $\epsilon\mu$  and  $\epsilon\alpha$  baits (Spearman's  $\rho = 0.70$ , Supplementary Fig. 1b), and was further reproduced in resting wild-type and activated AID-deficient B cells (*Aid*<sup>-/-</sup>) (Spearman's  $\rho = 0.93$  (resting) and 0.94 (*Aid*<sup>-/-</sup>); Fig. 1b, lanes 1–3, and Supplementary Fig. 2b). Nearly identical profiles were observed in B cells homozygous for an anti-hen egg lysozyme VDJ knock-in (anti-HEL, Spearman's  $\rho = 0.89$ , Supplementary Table 2), where most of the *Igh* variable domain is in germline configuration. Thus, globally, *Igh* nuclear interactions in peripheral B cells are largely independent of cell activation, AID expression, or *Ig* variable and constant region gene recombination.

4C-seq was validated by three-dimensional DNA fluorescence *in situ* hybridization (3D DNA FISH) using Perkin Elmer's ultra-high-throughput imaging system. This new approach allowed the automated and unbiased screening of 48,162 activated B cells. Analysis of *Igh* interactions with 14 genomic sites showed 3D FISH measurements to be in good agreement with 4C-seq ( $R^2 = 0.99$ , Supplementary Fig. 3).

### Genome features enriched in contacting loci

Even though the *Igh* and *Myc* loci are on different chromosomes their interactome was significantly correlated (Spearman's  $\rho = 0.58$  ( $P < 1 \times 10^{-8}$ ); Fig. 1b, lanes 1–3 and 5). This finding is consistent with the notion that these genes frequently associate and thus may share a common subnuclear environment in B cells<sup>15–17</sup>. To characterize

the genomic properties of loci interacting with *Igh* and *Myc*, we compared their 4C-seq profiles to genome-wide epigenetic and transcription maps<sup>13,18</sup>. The analyses revealed a good concordance between *Igh*- and *Myc*-interacting loci and activating chromatin acetylation, RNA polymerase II (Pol II) and messenger RNA transcripts (Fig. 1b, lanes 1–7, and Supplementary Fig. 2c, d). This correlation was particularly evident when entire chromosomes were considered. For example, *Igh* contact probability and Pol II were highest for chromosomes 11, 17 and 19, and lowest for chromosomes 3, 14 and 18 (Pearson's  $r = 0.92$  ( $P = 1.7 \times 10^{-8}$ ); Fig. 1c). This hierarchical ordering closely followed gene density estimates, which are highest (2.1%) for mouse chromosome 11, and lowest (1%) for chromosomes 3 and 14 (Supplementary Table 3). Similar correlations were obtained for *Myc*, although the correlations were lower than for *Igh* (Pearson's  $r = 0.61$  ( $P = 0.0013$ ); Supplementary Fig. 2d, e and Supplementary Table 4). Altogether, the data recapitulate the spatial compartmentalization of transcriptionally active, gene-dense domains<sup>6,11,14</sup>. In marked contrast, the interactome of transcriptionally silent *Mycn* in B cells or *Igh* in MEFs seemed to be random and did not correlate well with any of the genomic features surveyed, with the exception of centromeric regions for *Mycn* (Fig. 1b, lanes 8 and 9, and Supplementary Fig. 4). This latter feature might reflect the tendency of some silent loci to co-localize with peri-centromeric, repressive heterochromatin<sup>19</sup>. We conclude that in peripheral B cells *Igh* and *Myc* are more closely associated with transcribed, epigenetically accessible genomic sites, whereas interactions of transcriptionally inactive *Mycn* (or *Igh* in MEFs) are more randomly distributed.

### Translocations in the absence of AID

To examine the role of nuclear contacts on the genesis of chromosomal translocations in the absence of programmed DNA damage, we



compared the 4C-seq genomic profiles to translocation-capture sequencing (TC-seq) data sets obtained from *Aid*<sup>-/-</sup> activated B cells<sup>9</sup>. The TC-seq assay was recently developed to map genomic rearrangements comprehensively in B cells where a specific DNA break at *Igu* (also called *Ighm*) or *Myc* is created via expression of the I-SceI mega-nuclease<sup>9,12</sup>. In the absence of AID, *Igh*<sup>I-SceI</sup> or *Myc*<sup>I-SceI</sup> translocate to loci that occasionally suffer DNA damage as a result of normal metabolic processes such as transcription or DNA replication.

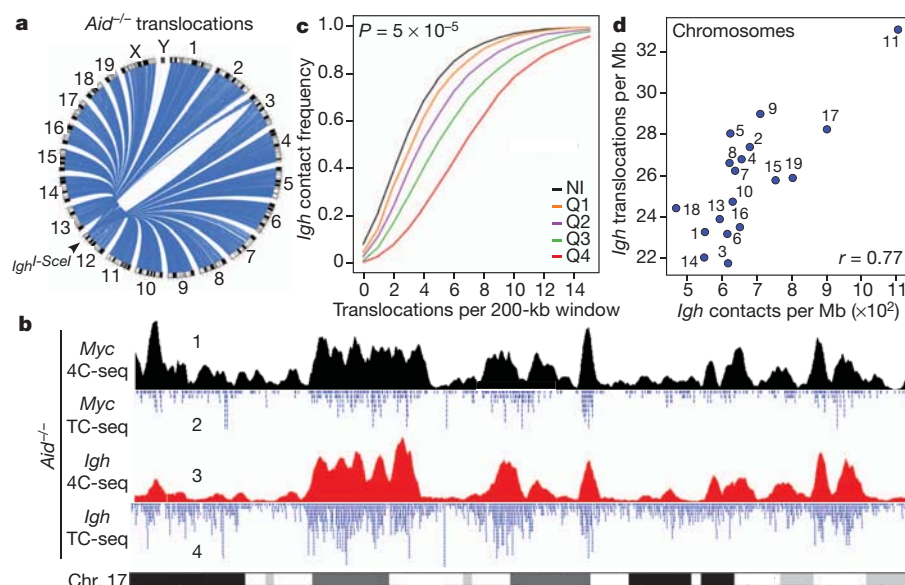
A total of 68,403 and 28,548 rearrangements were captured between *Igh*<sup>I-SceI</sup> and *Myc*<sup>I-SceI</sup>, respectively, and the rest of the genome (Fig. 2a; see also ref. 9). Visual comparison of the aligned 4C- and TC-seq reads revealed a nonrandom distribution of AID-independent rearrangements across chromosomes (Fig. 2b). Notably, the translocation profiles resembled the *Igh* and *Myc* interactome as well as accessible chromatin as measured by histone acetylation (Fig. 2b and Supplementary Fig. 5). Conversely, there was no obvious concordance between *Mycn* nuclear contacts in the same cells and *Igh*<sup>I-SceI</sup> or *Myc*<sup>I-SceI</sup> translocations (Supplementary Fig. 5). To validate these observations genome-wide, *Igh*, *Myc* and *Mycn* nuclear contacts were subdivided into quartiles (Q) and the data plotted as a function of total *Igh* or *Myc* translocations per 200-kb non-overlapping windows. The results showed that the greater the interaction frequency between *Igh* (or *Myc*) and a given genomic site, the more likely that the two loci were translocated (Q1 versus Q4,  $P = 0.0005$  (permutation test); Fig. 2c and Supplementary Fig. 6). In the case of *Igh*, where the number of captured rearrangements was substantial, translocations per chromosome were directly proportional to the contact frequency between *Igh* and a given chromosome (Pearson's  $r = 0.77$  ( $P = 0.0002$ ); Fig. 2d). Conversely, we observed little or no correspondence between *Igh* or *Myc* translocations and the interactome of *Mycn* in B cells (Q1 versus Q4,  $P = 0.35$ ; Supplementary Fig. 6). The data are thus consistent with the notion that AID-independent translocations occur preferentially between interacting genomic loci that are epigenetically accessible.

### AID-targeted translocations

AID produces lesions in a large number of defined hotspots, many of which are recurrent translocation partners for *Igh* in lymphoma<sup>2,9,10</sup>

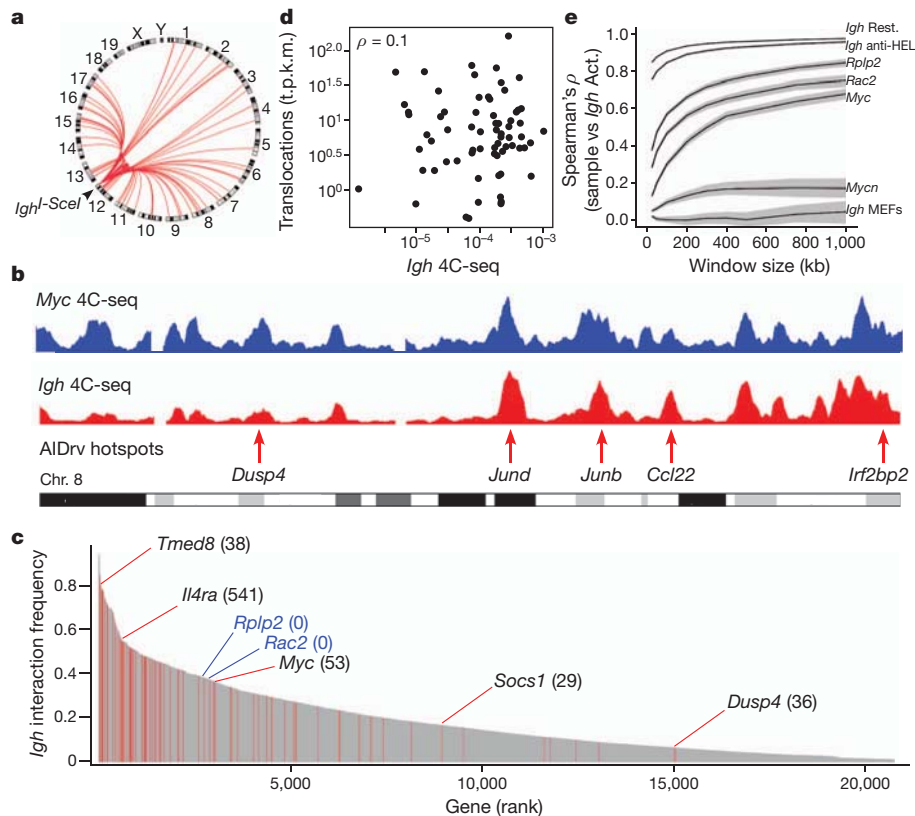
(Fig. 3a). To determine whether the location of these hotspots could be explained by B-cell nuclear architecture, we ranked RefSeq genes on the basis of 4C values. The analysis showed that a large fraction of loci carrying *Igh* translocation hotspots engaged in recurrent long-range interactions with this locus in activated B cells (Fig. 3b, c). At the same time, we identified thousands of genes that interacted repeatedly with *Igh* but that were not associated with translocation hotspots (Fig. 3c). For instance, up to 2,361 genes (11% of all RefSeq genes) outranked *Myc* in *Igh* contact frequency (Supplementary Table 5 and Supplementary Fig. 7), even though only 58 of them were recurrently rearranged to *Igh* (Fig. 3c). In a similar manner, whereas translocation hotspots in *Myc*<sup>I-SceI</sup> B cells were biased for domains co-localizing with *Myc*, physical proximity per se could not predict the presence of translocation hotspots in these cells (Supplementary Fig. 8). Furthermore, a subset of hotspot genes (for example, *Socs1* or *Dusp4*) associated infrequently with *Igh* (Fig. 3b, c), and we found no direct correlation between the number of translocations per hotspot and contact frequency with *Igh* (Spearman's  $\rho = 0.1$  ( $P = 0.4$ ); Fig. 3d). As an example, *Tmed8* and *Dusp4* genes were rearranged to *Igh* at roughly equal proportions (38 versus 36 translocations respectively), in spite of the fact that *Tmed8* was physically associated with *Igh* ~10 times more frequently than *Dusp4* in the B cell nucleus (Fig. 3c).

To exclude formally the possibility that *Igh* and *Myc* share translocation targets primarily because of shared contacts, we generated two additional 4C-seq libraries using baits specific for *Rac2* (chromosome 15) and *Rplp2* (chromosome 7). These genes are highly transcribed in B cells<sup>13</sup>, and both interact more frequently with *Igh* than *Myc* does, but neither is associated with AID-mediated translocation hotspots (Fig. 3c; see also ref. 9). The interactome of these two genes was then compared to that of *Igh* in activated B cells. As controls for this analysis, we included the interactomes of *Igh* from resting B cells, anti-HEL knock-in B cells and MEFs, as well as the *Mycn* interactome from stimulated B cells. As expected, the *Igh* interactome was similar in all B-cell types, but not in MEFs where it is not transcribed (Fig. 3e). Similarly, *Mycn*, which is transcriptionally silent in activated B cells, shows 4C-seq profiles with little correlation to *Igh* (Fig. 3e). Notably, the interactomes of *Rac2* and *Rplp2* were significantly more correlated to *Igh* than was *Myc* ( $P < 1 \times 10^{-5}$  (bootstrapping test);



**Figure 2 | Genomic distribution of AID-independent translocations correlates with nuclear contact profiles.** **a**, Genome-wide view of rearrangements to *Igh*<sup>I-SceI</sup> in *Aid*<sup>-/-</sup> B cells. **b**, Cross-comparison of contacts and translocations between *Myc* or *Igh* and mouse chromosome 17 in activated *Aid*<sup>-/-</sup> B cells. **c**, Empirical cumulative distribution showing *Igh*<sup>I-SceI</sup> *Aid*<sup>-/-</sup>

translocations per 200-kb non-overlapping windows as a function of *Igh* 4C-seq data subdivided as quartiles (Q1–4). NI represents windows with no aligned reads. **d**, Comparison of AID-independent translocations versus *Igh* 4C-seq per chromosome per mappable megabase. The degree of correlation is represented by Pearson's  $r$ .



**Figure 3 | Lack of correlation between translocation hotspots and nuclear architecture.** **a**, Genome-wide hotspots in activated B cells transduced with I-SceI and AID retroviruses (rv). **b**, Translocation hotspots (bottom lane) and *Myc* and *Igh* contacts in chromosome 8. **c**, *Igh* contacts with RefSeq genes. Hotspot<sup>+</sup> genes are highlighted in red, and for a subset of them the number of translocations is provided in parentheses. The two hotspot<sup>-</sup> genes (*Rplp2* and

Fig. 3e). Thus, nuclear interactions alone do not predispose transcriptionally active genes to high levels of translocation with *Igh*.

### A genome-wide map of AID-mediated dsDNA breaks

Our observations challenge the current view that preferential chromosome and/or gene locus interactions govern tumour-inducing translocations in AID-expressing B cells<sup>15,20</sup>, and indicate that the amount of AID-mediated DNA damage could account for the frequency of these events. To explore this idea we created a genome-wide map of AID-mediated DNA damage in activated B cells by measuring recruitment of replication protein A<sup>13,21</sup> (RPA) (Fig. 4a, lane 2). Because RPA accumulation is partially blocked by 53BP1 (also called Trp53bp1; refs 22, 23), we reasoned that genetic deletion of 53BP1 in B cells might increase RPA recruitment, thus providing a more sensitive means to map sites of AID-induced lesions by ChIP-seq. Consistent with this idea, RPA accumulation at *Igh* was markedly increased (7.8-fold relative to control) in the absence of 53BP1 (Fig. 4a, lane 3), and an even higher RPA signal (11-fold) was observed in 53BP1<sup>-/-</sup> mice overexpressing AID (*IgkAID*<sup>12</sup>; Fig. 4a, lane 4). Conversely, there was no detectable accumulation of RPA at *Igh* in activated *Aid*<sup>-/-</sup> 53BP1<sup>-/-</sup> B cells (Fig. 4a, lane 1). Thus, RPA recruitment to *Igh* is AID dependent and is enhanced in the absence of 53BP1.

In agreement with our previous findings<sup>13</sup>, we did not detect RPA recruitment at AID targets outside the *Igh* locus, such as *Cd83* (Fig. 4b, lane 2). However, we found prominent RPA ChIP signal at the same locus upon 53BP1 deletion (Fig. 4b, lanes 3 and 4). Analogous to *Igu* and *Igh* (also called *Ighg1*), the *Cd83* RPA island extended nearly 50 kb upstream and downstream of the transcription start site (TSS) (Fig. 4a, b, lane 4). In total, 153 non-*Ig* genes accumulated RPA in an AID-dependent fashion (Fig. 4c and Supplementary Table 6)

*Rac2*) highlighted in blue are discussed in more detailed in panel e. **d**, Scatter plot showing *Igh* translocations per hotspot versus contacts. Data are plotted as sequence tags per kb per million sequences (t.p.k.m.). **e**, Line graph showing the 4C-seq correlation (Spearman's  $\rho$ ) between *Igh* (in various cell types), *Rplp2*, *Rac2*, *Myc* and *Mycn* (in activated B cells) versus *Igh* in activated B cells. The 99% bootstrapping confidence intervals are shown in grey.

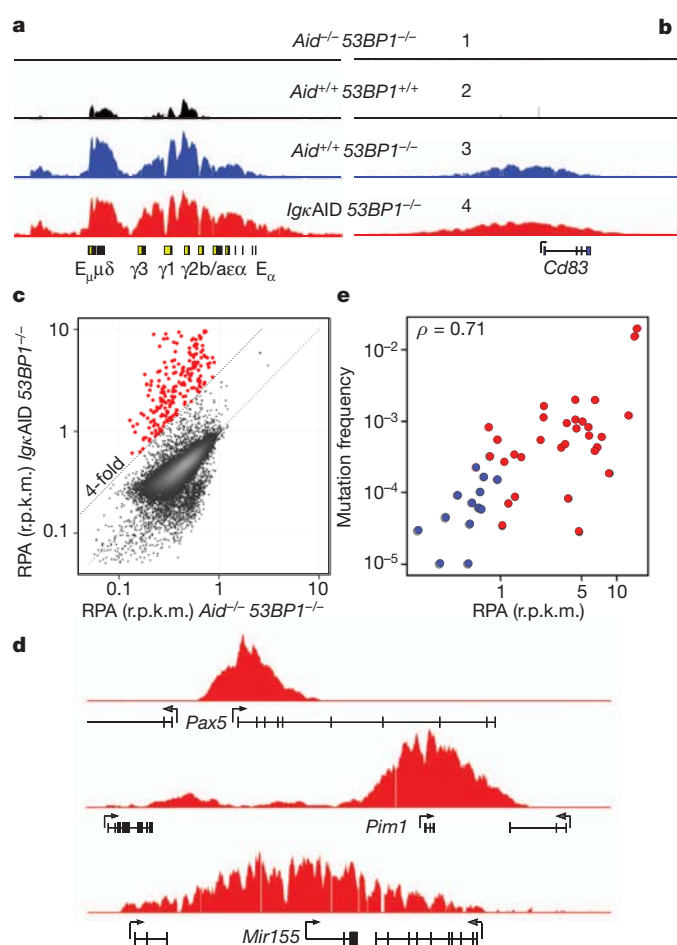
including known *Igh* translocation partners such as *Pax5*, *Pim1* and *Mir155* (Fig. 4d and Supplementary Table 6).

To ascertain the precise nature of RPA islands, we measured somatic hypermutation at a subset of RPA<sup>+</sup> and RPA<sup>-</sup> genes<sup>13</sup>. We found a strong positive correlation between the rate of hypermutation and the extent of RPA recruitment (Spearman's  $\rho = 0.71$ ; Fig. 4e and Supplementary Table 7). We conclude that RPA-seq can be used as a surrogate to measure AID-mediated DNA damage across the B-cell genome.

### Recurrent translocations are proportional to DNA damage

To evaluate the relative contribution of DNA damage to targeted translocations we compared the results of RPA-seq and TC-seq obtained from AID-expressing cells. We found a substantial overlap between the two data sets: out of a total of 97 genes with translocation hotspots with an average of 60 translocations per gene, 78 showed RPA accumulation (Supplementary Fig. 9 and Supplementary Table 8). A second group of genes (75) was also associated with RPA islands but displayed fewer translocations (mean = 7, Supplementary Fig. 9), and thus fell below our hotspot criteria cutoff. This result indicates that TC-seq underestimates the number of AID-mediated translocations, possibly due to lack of saturation<sup>9</sup>. Only 19 genes associated with translocation hotspots did not recruit RPA above background levels (Supplementary Fig. 9), suggesting that the RPA-seq data set is also not fully saturated. Thus, RPA demarcates sites of recurrent translocations in B lymphocytes.

In addition to the qualitative correlation above, we found that the absolute number of *Igh* translocations per hotspot was directly proportional to RPA recruitment (Spearman's  $\rho = 0.6$  ( $P = 2.9 \times 10^{-6}$ );

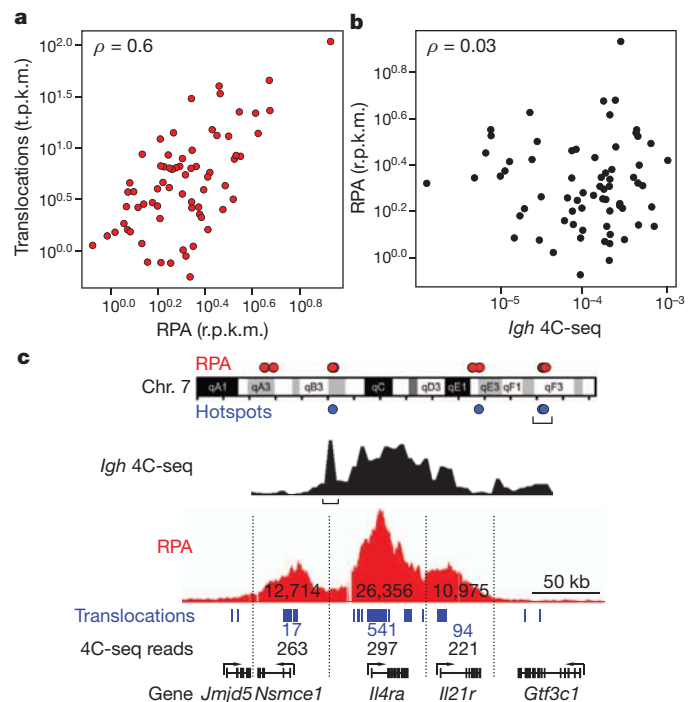


**Figure 4 | Genome-wide map of AID-mediated DNA damage.** **a**, RPA occupancy at *Igh* in activated B cells. Genotype and RPA sequence reads per million values are shown. **b**, Same analysis as in panel **a** for *Cd83*. For all tracks, background sequencing was filtered out via a threshold. **c**, One-hundred and fifty-three RPA islands (red dots) detected in *IgkAID 53BP1*<sup>-/-</sup> B cells fourfold above background (measured in *Aid*<sup>-/-</sup> *53BP1*<sup>-/-</sup> cells). Data are plotted as reads per kb per million sequences (r.p.k.m.). **d**, RPA islands associated with TSSs from *Pax5*, *Pim1* and *Mir155*. **e**, Hypermutation frequency relative to RPA recruitment at TSSs ( $\pm 2$  kb) in a subset of RPA<sup>+</sup> (red dots) and RPA<sup>-</sup> (blue dots) genes. Spearman's  $\rho$  is provided.

Fig. 5a). This result contrasts with the lack of correlation observed between nuclear contacts and total rearrangements per hotspot (Fig. 3d) or RPA accumulation (Spearman's  $\rho = 0.03$  ( $P = 0.8$ ); Fig. 5b). Similar results were observed for *Myc* (Spearman's  $\rho < 0.07$ ; data not shown). These findings demonstrate that with regard to AID-targeted translocations, DNA damage is the primary determinant of rearrangement location and frequency. This was particularly evident for AID targets that are clustered within  $\sim 200$ -kb genomic domains, such as *Nsmce1*, *Il4ra* and *Il21r* in chromosome 7 (Fig. 5c), or the *Hist1h1c* gene family in chromosome 13 (Supplementary Fig. 10 and Supplementary Table 9). Whereas we found little variation in *Igh* (or *Myc*) proximity for different genes within these clusters, translocations varied substantially and in a manner that was proportional to AID-mediated damage (Fig. 5c, Supplementary Fig. 10 and Supplementary Table 9). Taken together, these results clearly demonstrate that DNA break formation, but not nuclear interactions, governs the rate of recurrent chromosomal translocations.

## Discussion

We have shown that in the absence of programmed DNA damage, translocation partner selection is largely dictated by physical proximity, following principles of nuclear organization, chromatin accessibility



**Figure 5 | AID activity predicts the location and frequency of targeted chromosomal translocations.** **a**, Scatter plot showing the correlation between *Igh* translocations per hotspot and RPA recruitment. **b**, Same as panel **a** but *Igh* contact frequency is used instead of translocations. **c**, Upper schematic: distribution of RPA islands (red dots) and translocation hotspots (blue dots) in chromosome 7. Middle: *Igh* 4C-seq profile demarcating the *Nsmce1*-*Il4ra*-*Il21r* loci. Bottom: RPA islands, translocations and contact frequency for each gene.

and gene expression. Physical proximity has also been suggested to have an impact on the formation of recurrent translocations<sup>20</sup>. For instance, rearrangements between *BCR-ABL1* in chronic myeloid leukaemia, *RET-CDC6* in thyroid malignancies, *TMPRSS2-ERG/ETV1* in prostate cancer, and *PML-RARA* in acute promyelocytic leukaemia have all been ascribed to preferential interactions between translocating partners in tumour cell precursors<sup>4</sup>. Similarly, *Igh* and *Myc* chromosomes have been shown to associate in mouse and human B cells<sup>16,17</sup>, and RNA FISH has shown that the *Myc* and *Igh* alleles are frequently found in the same RNA Pol-II-enriched transcription factories<sup>15</sup>. Bystander translocations between *Igh* and *Igλ* have also been proposed to result from frequent contacts, as determined by 3D DNA FISH<sup>24</sup>. One limitation of FISH technology however is that it can only monitor a limited number of loci simultaneously. Consequently, it has been difficult to ascertain whether the documented contacts are truly unique relative to the broad array of genomic interactions. Our 4C measurements now clarify this issue in that they show that 29% of all genes interact with *Igh* at equal or higher frequency than *Myc*, *Igλ1*, or many of the oncogenes frequently rearranged in B-cell tumours. Thus, the rate of interaction between *Igh* and its recurrent translocation partners is not a specific feature of these loci and cannot account for their high rate of translocation.

Translocation requires joining of two double-stranded breaks (DSBs). Therefore, when breaks are limiting, increasing their frequency increases the rate of translocation<sup>24,25</sup>. Similarly, repair deficiencies augment the rate of translocation by increasing the half-life of dsDNA breaks and thereby the availability of substrates for aberrant repair<sup>2</sup>. However, it has not been possible to relate directly the frequencies of DNA damage and translocations because neither the extent nor the location of DSBs in the B-cell genome was known. We have overcome these limitations by measuring RPA deposition at sites of DNA damage in 53BP1 mutant B cells. On the basis of this new strategy we uncovered  $\sim 150$  non-*Ig* genes that suffer AID-mediated DSBs. These genes



coincide with translocation hotspots, and are sites of ongoing hypermutation. Most importantly, by relating RPA-seq with TC-seq and 4C-seq data sets we found that the frequency of DNA damage directly accounts for the rate of translocation, as shown by the marked concordance between the amount of RPA deposition and the absolute number of rearrangements at any given genomic site. This view is also supported by the relative lack of correlation between proximity and the absolute number of rearrangements per hotspot.

The genomic distribution of sporadic translocations is best explained by nuclear architecture, whereas the location and incidence of recurrent translocations, including those involved in B-cell malignancies, directly reflect site-specific DNA damage.

## METHODS SUMMARY

Full details of B-cell culture, hypermutation analysis, chromatin immunoprecipitation, chromosome conformation capture on Chip (4C), translocation capture sequencing analysis, deep sequencing and bioinformatics techniques are provided in Methods. The NIAMS-NIH Animal Care and Use Committee approved all animal protocols and experiments.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 11 August 2011; accepted 31 January 2012.**

**Published online 7 February 2012.**

- Mitelman, F., Johansson, B. & Mertens, F. The impact of translocations and gene fusions on cancer causation. *Nature Rev. Cancer* **7**, 233–245 (2007).
- Nussenzweig, A. & Nussenzweig, M. C. Origin of chromosomal translocations in lymphoid cancer. *Cell* **141**, 27–38 (2010).
- Tsai, A. G. *et al.* Human chromosomal translocations at CpG sites and a theoretical basis for their lineage and stage specificity. *Cell* **135**, 1130–1142 (2008).
- Zhang, Y. *et al.* The role of mechanistic factors in promoting chromosomal translocations found in lymphoid and other cancers. *Adv. Immunol.* **106**, 93–133 (2010).
- Cremer, T. & Cremer, M. Chromosome territories. *Cold Spring Harb. Perspect. Biol.* **2**, a003889 (2010).
- Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
- Hakim, O., Sung, M. H. & Hager, G. L. 3D shortcuts to gene regulation. *Curr. Opin. Cell Biol.* **22**, 305–313 (2010).
- Chakalova, L. & Fraser, P. Organization of transcription. *Cold Spring Harb. Perspect. Biol.* **2**, a000729 (2010).
- Klein, I. A. *et al.* Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. *Cell* **147**, 95–106 (2011).
- Chiarle, R. *et al.* Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. *Cell* **147**, 107–119 (2011).
- Simonis, M. *et al.* Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nature Genet.* **38**, 1348–1354 (2006).
- Robbiani, D. F. *et al.* AID produces DNA double-strand breaks in non-Ig genes and mature B cell lymphomas with reciprocal chromosome translocations. *Mol. Cell* **36**, 631–641 (2009).
- Yamane, A. *et al.* Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. *Nature Immunol.* **12**, 62–69 (2011).
- Hakim, O. *et al.* Diverse gene reprogramming events occur in the same spatial clusters of distal regulatory elements. *Genome Res.* **21**, 697–706 (2011).
- Osborne, C. S. *et al.* Myc dynamically and preferentially relocates to a transcription factory occupied by Igh. *PLoS Biol.* **5**, e192 (2007).
- Roix, J. J., McQueen, P. G., Munson, P. J., Parada, L. A. & Misteli, T. Spatial proximity of translocation-prone gene loci in human lymphomas. *Nature Genet.* **34**, 287–291 (2003).
- Parada, L. A., McQueen, P. G. & Misteli, T. Tissue-specific spatial organization of genomes. *Genome Biol.* **5**, R44 (2004).
- Kuchen, S. *et al.* Regulation of microRNA expression and abundance during lymphopoiesis. *Immunity* **32**, 828–839 (2010).
- Brown, K. E., Baxter, J., Graf, D., Merkenschlager, M. & Fisher, A. G. Dynamic repositioning of genes in the nucleus of lymphocytes preparing for cell division. *Mol. Cell* **3**, 207–217 (1999).
- Meaburn, K. J. & Misteli, T. Cell biology: chromosome territories. *Nature* **445**, 379–381 (2007).
- Vuong, B. Q. *et al.* Specific recruitment of protein kinase A to the immunoglobulin locus regulates class-switch recombination. *Nature Immunol.* **10**, 420–426 (2009).
- Bunting, S. F. *et al.* 53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks. *Cell* **141**, 243–254 (2010).
- Bothmer, A. *et al.* 53BP1 regulates DNA resection and the choice between classical and alternative end joining during class switch recombination. *J. Exp. Med.* **207**, 855–865 (2010).
- Wang, J. H. *et al.* Mechanisms promoting translocations in editing and switching peripheral B cells. *Nature* **460**, 231–236 (2009).
- Robbiani, D. F. *et al.* AID is required for the chromosomal translocations in *c-myc* that lead to *c-myc/IgH* translocations. *Cell* **135**, 1028–1038 (2008).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank members of the Casellas and Nussenzweig laboratories for discussions; G. Gutierrez from NIAMS genomics facility for technical assistance. This work was supported in part by a grant from the Starr Foundation to M.C.N., by NIH grant number AI037526 to M.C.N. and the Intramural Research Program of NIAMS and NCI, NIH. M.C.N. is an HHMI investigator. This study made use of the high-performance computational capabilities of the Biowulf Linux cluster at the NIH (<http://biowulf.nih.gov>), and the resources of NCI's High-Throughput Imaging Facility.

**Author Contributions** R.C., O.H., G.L.H. and M.C.N. planned studies and interpreted data. Experiments were performed as follows: O.H. and C.A.-S., 4C-seq and FISH; A.Y., RPA-seq; I.K., A.B., D.F.R. and M.J., TC-seq; W.R., E.M. and T.O., bioinformatics; K.-R.K.-K., T.C.V., H.N. and J.C., FISH; G.L. and H.N., hypermutation; A.N., 53BP1 expertise; M.C.N. and R.C. wrote the manuscript.

**Author Information** All sequence data are available at the NCBI SRA database under accession number SRP010565. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to R.C. ([casellar@mail.nih.gov](mailto:casellar@mail.nih.gov)) or M.C.N. ([nussen@rockefeller.edu](mailto:nussen@rockefeller.edu)).

## METHODS

**B-cell activation and hypermutation analysis.** Miltenyi microbead-isolated CD43<sup>+</sup> splenic B cells from wild-type, *IgkAID-Ung*<sup>-/-</sup>, or *Aicda*<sup>-/-</sup> mice were cultured at  $0.1 \times 10^6$  cells per ml with  $50 \mu\text{g ml}^{-1}$  lipopolysaccharide (LPS) (Sigma),  $2.5 \text{ ng ml}^{-1}$  mouse recombinant IL-4 (Sigma) and  $0.5 \mu\text{g ml}^{-1}$  of aCD180 (RP105) antibody (RP14, BD Pharmingen). For 4C-seq and TC-seq procedures, cells were collected at 72 h. For hypermutation analysis, cells were diluted 1:4 at 72 h and cultured for another 48 h under the same conditions. Fifty nanograms of genomic DNA was then amplified for 30 cycles with Phusion DNA polymerase and gene-specific primers. When nested PCR was applied, 40 (20 + 20) cycle amplifications were performed in the presence of DMSO. The amplicon was cloned using PCR Zero blunt (Invitrogen) and sequenced.

**Chromosome conformation capture on chip (4C) followed by deep-sequencing.** The 4C assay was performed as previously described<sup>14</sup> with minor modifications. Ten million mouse B cells were crosslinked in 2% formaldehyde at 37 °C for 10 min. The reaction was quenched by the addition of glycine (final concentration of 0.125 M). Cells were then washed with cold PBS and lysed (10 mM Tris-HCl, pH 8.0, 10 mM NaCl, 0.2% NP-40, 1× complete protease inhibitors (Roche)) at 4 °C for 1 h. Nuclei were incubated at 37 °C for 1 h in 500  $\mu\text{l}$  of restriction buffer (New England Biolabs buffer 2 for HindIII or buffer 3 for BglII digestion) containing 0.3% SDS. To sequester SDS, Triton X-100 was then added to a final concentration of 1.8%. DNA digestion was performed with 400 U of HindIII or BglII (New England Biolabs) at 37 °C overnight. After heat inactivation (65 °C for 30 min), the reaction was diluted to a final volume of 7 ml with ligation buffer containing 100 U T4 DNA Ligase (Roche) and incubated at 16 °C overnight. Samples were then treated with 500  $\mu\text{g}$  Proteinase K (Ambion) and incubated overnight at 65 °C to reverse formaldehyde crosslinking. DNA was then purified by phenol extraction and ethanol precipitation. For circularization, the ligation junctions were digested with Csp6I (Fermentas) or DpnII (New England Biolabs) at 37 °C overnight. After enzyme inactivation and phenol extraction, the DNA was religated in a 7-ml volume (1,000 U T4 DNA Ligase, Roche). Three micrograms of 4C library DNA was amplified with Expand Long template PCR System (Roche). Thermal cycle conditions were DNA denaturing for 2 min at 94 °C, followed by 30 cycles of 15 s at 94 °C, 1 min at 60 °C, 3 min at 68 °C, and a final step of 7 min at 68 °C. Baits were amplified with inverse PCR primers as follows: *Igh* with HindIII: IgH\_R\_4C 5'-CCAGACATGTGG GCTGAGAT-3', *Igh*\_Hind\_Read 5'-CTACCCACCTAACTCCAAGC-3'; *Mycn* with HindIII: Mycn\_R\_4C 5'-CTCCCATTTTGCACTCTCTGT-3', *Mycn*\_Hind\_Read 5'-GATTTATCCTTAAACCTTAAGC-3'; *Igh* with BglII: IgH\_Bgl\_R\_4C 5'-CATGGACATTTGCGTGTGTA-3', *IgH*\_Bgl\_Read 5'-GTG CCCCCAGGAGCAGATCT-3'; *Mycn* with BglII: Mycn\_Bgl\_R\_4C 5'-AG TCTCGGGAGGTAAGAAG-3', *Mycn*\_Bgl\_Read 5'-CCCTTTAGACAGCC AGATCT-3'; *Myc* with BglII: Myc\_Bgl\_R\_4C 5'-AAGAATGTGCCAGTC AACA-3', *Myc*\_Bgl\_Read 5'-AGTGAATTGCCAACCAGAT-3'; *Rplp2* with HindIII: 5'-GCCATCTCTCCAGTCAAAAAGC-3', *CTTCTCACTTCCATT CCCTGAG-3'*; *Rac2* with HindIII: 5'-GCCATGGAGACCGGAAGCTT-3', 5'-GGGACTGTCCACTCCACTT-3'; *Eμ* with HindIII: 5'-TGTGGCTGCTGC TCTTAAAGC-3', 5'-TGTGAAGCCGTTTGTACCAGAATGT-3'. 4C amplified DNA was microsequenced with the Illumina platform. For multiplexing purposes, extra nucleotides were added at the 5' end of read primers: a T for LPS + IL-4 activated B cells, TT for resting B cells, A for IgH<sup>2HEL/2HEL</sup> B cells, and AA for MEFs.

**Translocation capture sequencing.** All experimental procedures involving translocation capture sequencing (TC-seq) library preparation and computational analysis is provided in ref. 9.

**Bioinformatics.** For 4C-seq, standard Illumina pipeline software (version  $\geq 1.8$ ) was used to process raw data and obtain Fastq files of paired short reads. Each read pair was then tested for the presence of a perfect match to the respective bait primer as well as the bait spacer between the end of the primer and the restriction sites used in the corresponding experiment. In some experiments up to three mismatches in the non-index portion of the PCR primer/flank sequence was allowed without any relevant changes in the resulting data. These flanking sequences other than the restriction site were trimmed and the remainder was aligned against the mouse genome (build mm9/NCBI37) with Bowtie<sup>26</sup> with the following command line options: '-X 500 -p 3 -v2 -k2 -m1-phred64-quals -sam', which reported all unique alignments with at most two mismatches. In the case of Eμ, only the HindIII spanning read in each pair was sufficiently long to reach the ligated interaction partner. In that case, single end alignments of the flank were carried out with the command line options '-best-all-strata-chunkmbs 256 -m1 -sam'. Alignments were then matched up with restriction sites and assigned to a HindIII or BglII fragment. Fragments were combined into 200-kb non-overlapping windows to determine (1) the

total number of 4C reads per fragment, and (2) the fraction of restriction fragments for which 4C reads were found. The latter part of the analysis was carried out with a combination of custom software written in Bash, Python, R, and BedTools<sup>27</sup>. Processing of fragment- or window-level data was carried out in R using standard methods.

For RPA-seq, short reads obtained from Illumina pipeline were aligned against the mouse genome (build mm9/NCBI37) with Bowtie<sup>26</sup> using options '-threads = 8-phred64-quals-best-all-strata -m1 -n2 -l36 -sam'. Raw tag densities of *IgkAID 53BP1*<sup>-/-</sup> and *Aid*<sup>-/-</sup> *53BP1*<sup>-/-</sup> samples flanking transcription start sites of RefSeq genes were compared and genes with >4-fold enrichment were selected.

For histone acetylation, short reads obtained for ChIP against a set of histone acetylations (see below) were aligned above. Areas of local enrichment over a random background model (islands) were identified with SICER 1.03<sup>28</sup> and the density of reads overlapping these islands was averaged in the same 200-kb non-overlapping windows used for 4C analysis.

**High-throughput 3D DNA FISH.** For 3D FISH, cultured B cells were set in a 384-well, poly-D-lysine-coated microplate (Perkin Elmer) by centrifugation at 1,000 r.p.m. for 5 min. After fixation in 4% PFA for 10 min, and permeabilization in 0.5% saponin (Sigma Aldrich)/0.5% Triton X-100/PBS for 20 min, cells were incubated in 0.1 N HCl for 10 min. Two PBS washes were applied between each step. After a 2× SSC wash, cells were kept in 50% formamide/2× SSC buffer for at least 30 min. Bacterial artificial chromosomes (BACs) were used as probes as follows: *Mlh3*, RP24-139J8; *Klhdc1*, RP24-109D18; *Clec2d*, RP24-149B3; *Mir142*, RP24-376D9, *Il4ra*, RP23-60A3; *Cytl1*, RP23-267B12; *Pim1*, RP24-331E7; *Furin*, RP24-377F13; *Pax5*, RP23-258E20; *Myc*, RP24-297E9; *Cxcr5*, RP24-308P6; *Rasa3*, RP24-247P3; *Gata3*, RP24-402N11. For each BAC, single colonies were grown and the presence of BAC DNA was verified by PCR. DNA was isolated and labelled with biotin (Roche; Biotin-Nick translation mix) or digoxigenin (Roche; DIG-Nick translation mix). A probe mix containing 250 ng of digoxigenin- and biotin-labelled probes, 3  $\mu\text{g}$  mouse COT1 DNA (Invitrogen), and 20  $\mu\text{g}$  tRNA (Ambion) was ethanol precipitated, and re-suspended in 15  $\mu\text{l}$  of hybridization buffer (10% dextran sulphate, 50% formamide, 2× SSC, and 1% Tween 20). Probe was then added to each well, denatured together with nuclei at 85 °C for 7 min and left to hybridize at 37 °C overnight in a humidified chamber. To discard non-hybridizing probe, cells were then washed in 50% formamide with 2× SSC at 45 °C, followed by washes with 1× SSC at 60 °C. Each wash was repeated three times for 5 min. Cells were blocked with 3% BSA/0.05% Tween 20/4× SSC for 20 min at room temperature and then incubated for 1 h with Fluorescein Avidin (Vector) and anti-Dig-Rhodamin (Roche) diluted 1:200 in blocking solution. Next, cells were washed three times with 0.05% Tween 20/4× SSC and mounted in DAPI-containing Vectashield (Vector) for imaging.

**Microscopy.** Cells were imaged in 384-well plates with the Opera (Perkin Elmer) confocal high-throughput imaging system using a ×40 water objective lens with 9 optical steps of 1.0  $\mu\text{m}$ .

**Automated image analysis algorithms.** To quantify distances between FISH signals, we customized the automated image analysis computer algorithm from ref. 14 to allow analysis of Opera images. This algorithm determines centre-to-centre distances of FISH signals in 3D. Nuclei and DNA FISH loci were automatically identified by a combination of the Acapella image analysis software (Perkin Elmer) together with a series of custom algorithms developed using Matlab technical computing software and the Matlab Image Processing toolbox (The Mathworks). The resulting morphometric information from each nucleus was automatically stored in a Matlab database file, which could be accessed by the custom algorithms. First, nuclei region of interest (nucROI) were segmented using the DAPI signal channel by the Acapella software. Then DNA FISH loci (fishROI) were automatically identified by a separate algorithm in the nucROI positional information. FishROIs were identified by intensity-based thresholding of the FISH fluorescent channel images. A third algorithm assigned the nuclear position of each FISH locus, based on the  $x$ - $y$ - $z$  location of the centre of the brightest  $3 \times 3$  FISH fluorescent channel pixel that was located within the fishROI. For each nucleus, three-dimensional ( $x$ - $y$ - $z$ ) distances were calculated between all possible pairs of FISH loci. The closest distance between the two probes in each nucleus was used to calculate the frequency of cells with distances <1  $\mu\text{m}$  as previously described<sup>14</sup>. Interaction frequencies between each of the genes and *Igh* was calculated by computing the percentage of cells carrying at least one pair of FISH signals separated by distances smaller than 1  $\mu\text{m}$  (Supplementary Fig. 12a). This distance threshold was recently shown to correlate well with 4C contact frequencies<sup>14</sup>. To determine the minimum number of cells necessary to reach statistical significance in our experimental setting, we first examined 17,462 LPS plus IL-4 activated lymphocytes with *Igh*- and *Myc*-specific probes. We found that at least 2,000

cells were necessary to reach a standard error of 0.01 (Supplementary Fig. 12b). For different probes, the optimal number of cells was as follows: 2,200 for *Myc*, 4,300 for *Mlh3*, 2,100 for *Clec2d*, 3,800 for *Klhdc1*, 1,500 for RP24-137A18, 1,300 for *Gata3*, 2,200 for *Il4ra*, 1,900 for *Mir142*, 2,200 for *Cyth1*, 2,100 for *Pim1*, 2,100 for *Furin*, 1,900 for *Pax5*, 1,800 for *Cxcr5* and 1,300 for *Rasa3*.

26. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
27. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
28. Zang, C. *et al.* A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* **25**, 1952–1958 (2009).



# RR-Lyrae-type pulsations from a 0.26-solar-mass star in a binary system

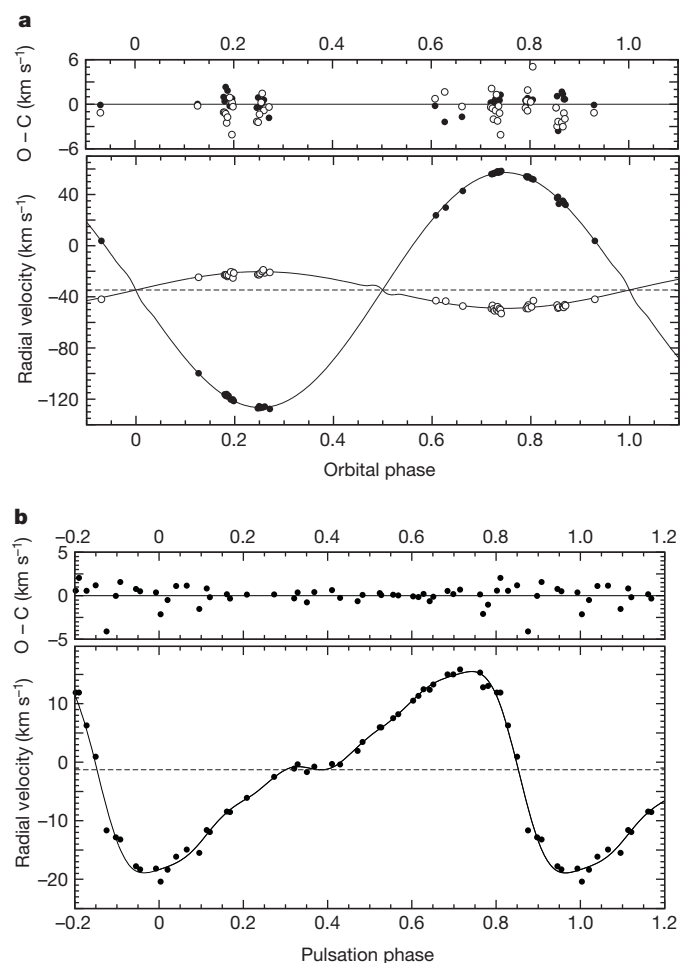
G. Pietrzyński<sup>1,2</sup>, I. B. Thompson<sup>3</sup>, W. Gieren<sup>1</sup>, D. Graczyk<sup>1</sup>, K. Stępień<sup>2</sup>, G. Bono<sup>4,5</sup>, P. G. Prada Moroni<sup>6,7</sup>, B. Pilecki<sup>1,2</sup>, A. Udalski<sup>2</sup>, I. Soszyński<sup>2</sup>, G. W. Preston<sup>3</sup>, N. Nardetto<sup>8</sup>, A. McWilliam<sup>3</sup>, I. U. Roederer<sup>3</sup>, M. Górski<sup>1,2</sup>, P. Konorski<sup>1,2</sup> & J. Storm<sup>9</sup>

RR Lyrae pulsating stars have been extensively used as tracers of old stellar populations for the purpose of determining the ages of galaxies, and as tools to measure distances to nearby galaxies<sup>1–3</sup>. There was accordingly considerable interest when the RR Lyrae star OGLE-BLG-RRLYR-02792 (referred to here as RRLYR-02792) was found to be a member of an eclipsing binary system<sup>4</sup>, because the mass of the pulsator (hitherto constrained only by models) could be unambiguously determined. Here we report that RRLYR-02792 has a mass of 0.26 solar masses ( $M_{\odot}$ ) and therefore cannot be a classical RR Lyrae star. Using models, we find that its properties are best explained by the evolution of a close binary system that started with  $1.4M_{\odot}$  and  $0.8M_{\odot}$  stars orbiting each other with an initial period of 2.9 days. Mass exchange over 5.4 billion years produced the observed system, which is now in a very short-lived phase where the physical properties of the pulsator happen to place it in the same instability strip of the Hertzsprung–Russell diagram as that occupied by RR Lyrae stars. We estimate that only 0.2 per cent of RR Lyrae stars may be contaminated by systems similar to this one, which implies that distances measured with RR Lyrae stars should not be significantly affected by these binary interlopers.

Using high-resolution spectra obtained with the MIKE spectrograph at the 6.5-m Magellan Clay telescope at the Las Campanas Observatory in Chile, and the UVES spectrograph attached to the 8.2-m VLT telescope of the European Southern Observatory on Paranal, we confirmed that RRLYR-02792 is a true physical, well detached, double-lined eclipsing binary system very well suited for deriving the masses of its two components with very high accuracy.

Analysis of the spectroscopic and photometric observations (Figs 1 and 2) results in the determination of the astrophysical parameters of our system presented in Supplementary Table 1. Realistic errors of the derived system parameters were determined using Monte Carlo simulations. The resulting masses of the components turned out to be very unexpected. The dynamical mass of the RR Lyrae component,  $(0.261 \pm 0.015)M_{\odot}$ , is much smaller than the mass required for helium ignition, and therefore completely at odds with the predictions of all theoretical models of RR Lyrae stars<sup>5–7</sup>. Moreover, if the pulsating component of RRLYR-02792 (star 1, temperature  $T_1$ ) were indeed a classical RR Lyrae star, as suggested by its light curve and pulsation period, the nature of the more massive, cooler (at  $T_1 = 6,000$  K,  $T_2 = 0.68 \times T_1$ ) and fainter (by some two magnitudes in the V band) secondary component (star 2) would be extremely unusual. Assuming a typical temperature for the RR Lyrae star of 6,000 K, the temperature of the static secondary component ( $T_2$ ) would be only about 4,100 K, much too cool for a giant star with  $M_2 = 1.67M_{\odot}$  (whose temperature is expected to be close to 5,000 K).

A clue comes from the relatively short orbital period of 15.24 days, which suggests that mass exchange between the two components should have occurred during the evolution of this system. Inspired

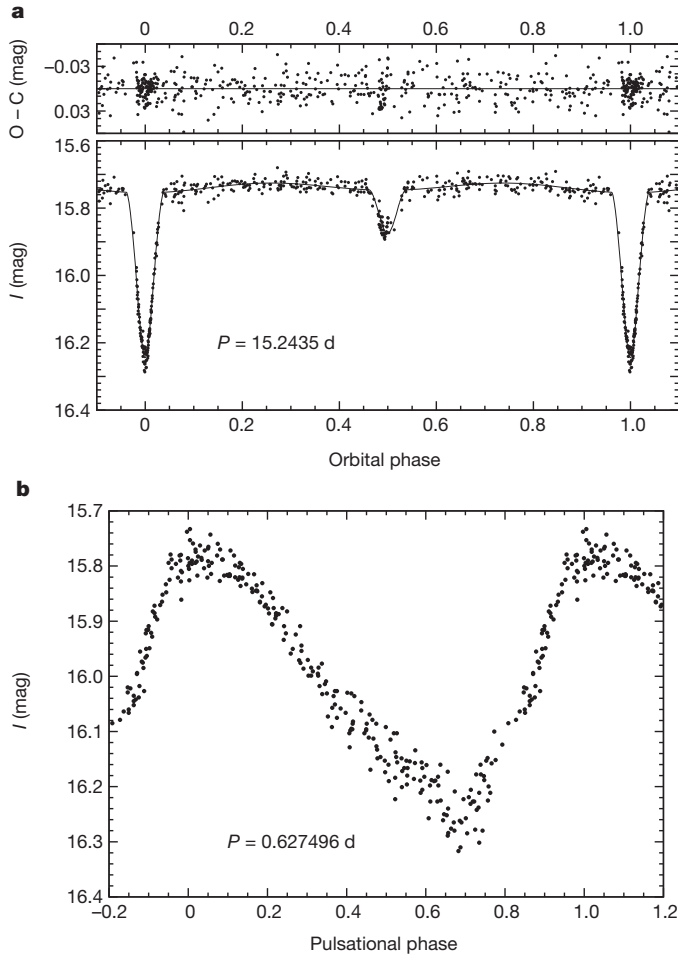


**Figure 1 | Orbital motion of the two binary components, and the pulsational motion of the pulsating component of the RRLYR-02792 system.** **a**, Main panel, the disentangled orbital radial velocity curves of both components of our binary system. Filled circles and open circles, primary and secondary component, respectively. **b**, The pulsational radial velocity curve of the primary component. The top panel in **a** and **b** shows the residuals of the fits (see below): observed radial velocities (O) minus the computed radial velocities (C). All individual radial velocities were determined by the cross-correlation method using appropriate template spectra and the MIKE and UVES spectra, yielding in all cases velocity accuracies better than  $300 \text{ m s}^{-1}$  (error bars smaller than the data symbols). Then, the orbit (mass ratio, systemic velocity, velocity amplitudes, eccentricity and periastron passage) plus a Fourier series of order eight approximating the pulsation variations of the primary component was fitted with a least squares method to the measured velocities. The resulting parameters are presented in Supplementary Table 1.

<sup>1</sup>Departamento de Astronomía, Universidad de Concepción, Casilla 160-C, Concepción, Chile. <sup>2</sup>Warsaw University Observatory, Al. Ujazdowskie 4, 00-478 Warsaw, Poland. <sup>3</sup>Carnegie Observatories, 813 Santa Barbara Street, Pasadena, California 91101-1292, USA. <sup>4</sup>Dipartimento di Fisica Università di Roma Tor Vergata, via della Ricerca Scientifica 1, 00133 Rome, Italy. <sup>5</sup>INAF-Osservatorio Astronomico di Roma, Via Frascati 33, 00040 Monte Porzio Catone, Italy. <sup>6</sup>Dipartimento di Fisica 'E. Fermi', Università di Pisa, Largo B. Pontecorvo, 3, I-56127, Pisa, Italy. <sup>7</sup>INFN, Sezione di Pisa, Largo B. Pontecorvo, 3, I-56127, Pisa, Italy. <sup>8</sup>Laboratoire Lagrange, UMR7293, UNS/CNRS/OCA, 06300 Nice, France. <sup>9</sup>Leibniz Institute for Astrophysics, An der Sternwarte 16, 14482 Potsdam, Germany.

by this possibility, we calculated a series of models for Algol-type binary systems<sup>8,9</sup>. We found that a system which initially contained two stars with  $M_1 = 1.4M_\odot$  and  $M_2 = 0.8M_\odot$  orbiting each other with an initial period of 2.9 days would, after 5.4 Gyr of evolution, have exchanged mass between the components as classical Algols do, and today would form a system very similar to RRLYR-02792 (with  $M_1 = 0.268M_\odot$  and  $M_2 = 1.665M_\odot$ , and orbital period  $P = 15.9$  days).

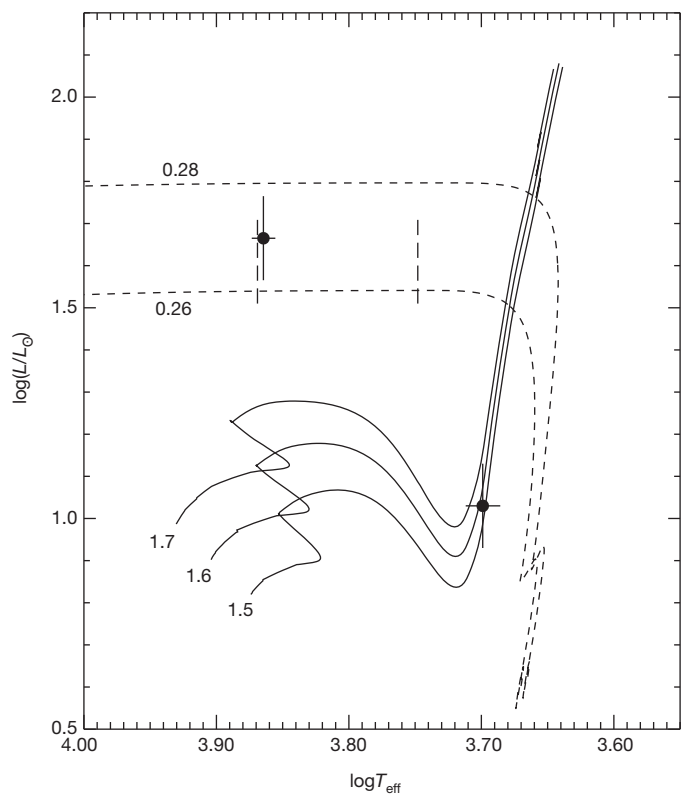
We therefore conclude that the primary component of our observed system is not a classical RR Lyrae star with its well-known internal structure. Rather, it is a star that possesses a partially degenerate helium core and a small hydrogen-rich envelope (undergoing shell burning) that has lost most of its envelope during the previous red giant branch



**Figure 2 | Change of brightness of the binary system caused by the mutual eclipses, and the intrinsic change of the brightness of the primary component caused by its pulsations.** **a**, Main panel, orbital I-band light curve (617 epochs collected over 10 years) of the binary system RRLYR-02792, after removal of the intrinsic brightness variation of the pulsating component (data points), together with the solution (solid line), as obtained with the 2007 version of the standard Wilson-Devinney code<sup>11,12</sup>. Top panel, the residuals of the observed magnitudes from the computed orbital light curve. **b**, Pulsational I-band light curve of the primary component of our binary system, folded on a pulsation period of 0.627496 days. The shape of the light curve is mimicking that of a classical RR Lyrae star. The following final ephemeris for our system was derived from the OGLE photometric data:  $P = 15.24340 \pm 0.00021$  d,  $T_0 = 2,452,108.3161 \pm 0.038$  d (orbital);  $P = 0.627496 \pm 0.000008$  d,  $T_0 = 2,455,000.355 \pm 0.005$  d (pulsation). Adopting the photometric ephemeris, and the mass ratio obtained from the analysis of the spectroscopic data (Fig. 1), we model our spectroscopic and photometric observations using the Wilson-Devinney code. We accounted for the intrinsic photometric variations of the pulsating star in the system by fitting a Fourier series of order 15 to the observations secured outside the eclipses and then subtract the corresponding variations in an iterative way, scaling the obtained fit according to the obtained Wilson-Devinney model.

phase to the secondary star due to mass exchange in the binary system. It is now evolving towards the hot subdwarf region on the Hertzsprung–Russell diagram (see Fig. 3 and Supplementary Fig. 2).

The pulsational light curve of such a star very closely resembles that of a classical RR Lyrae star. However, variable stars produced this way are expected to cross the classical pulsational instability region on the Hertzsprung–Russell diagram about 100 times faster than do the RR Lyrae stars. Because the star is moving rapidly at a constant luminosity across the instability strip towards higher temperatures, its radius should become smaller and therefore its pulsation period should steadily decrease. Indeed, using our photometric data we have measured a period decrease of the pulsating component of  $(8.4 \pm 2.6) \times 10^{-6}$  d  $\text{yr}^{-1}$ , which is on average more than two orders of magnitude larger than the period change shown by canonical RR Lyrae stars<sup>10</sup>, and therefore strongly supports our interpretation. Moreover, we have detected hydrogen lines in the spectrum of the binary associated with the pulsating primary component (see Supplementary Fig. 1), which



**Figure 3 | Positions of the two stars in the RRLYR-02792 binary system on the Hertzsprung–Russell diagram.** The two stars are shown by filled circles (error bars,  $1\sigma$ );  $L$  and  $L_\odot$ , stellar and solar luminosity, respectively;  $T_{\text{eff}}$ , stellar effective temperature. Solid lines show evolutionary models computed for 1.7, 1.6 and 1.5  $M_\odot$  stars with the most recent version of the FRANEC evolutionary code adopting updated input physics<sup>13</sup>. Current evolutionary models were computed assuming a solar chemical composition (metals,  $Z = 0.0129$ ; helium,  $Y = 0.274$ ). We adopted the recent heavy-element solar mixture<sup>14</sup> and a mixing-length value of  $\alpha = 1.74$ . Dashed lines proceeding horizontally from the luminosity axis show evolution of stellar structures with final masses of 0.26  $M_\odot$  and 0.28  $M_\odot$ , computed by following the standard evolution of a 1.4  $M_\odot$  stellar structure from the pre-main sequence up to the beginning of the red giant phase. At  $\log(L/L_\odot) = 0.9$  we applied an enhanced mass loss rate of about  $10^{-7} M_\odot \text{yr}^{-1}$  until the final masses (that is, 0.26  $M_\odot$  and 0.28  $M_\odot$ ) were approached. We computed the final evolutionary fate of these structures, at constant mass, down to the cooling phase of He-core white dwarfs<sup>15</sup>. The two short vertical dashed lines show the instability strip for typical RR Lyrae stars according to models for a solar chemical composition ( $Z = 0.02$ ,  $Y = 0.28$ )<sup>16</sup>, in which the pulsating component of our binary system is located. We adopted 300 K as the uncertainty of the calculated instability strip. Very good agreement between the evolutionary models and the observations is demonstrated.

confirms that the star possesses a hydrogen-rich envelope. In such a scenario, the secondary component is a typical red giant star currently evolving up the red giant branch, increasing its size and luminosity. During its future evolution, our system will turn into a binary system composed of two white dwarfs sharing a common envelope.

We have captured the RRLYR-02792 binary system in a very special and short-lived phase of its evolution, which constitutes just a small fraction ( $10^{-4}$ ) of its current age. The system provides a number of strong observational constraints that have enabled us to track its past evolution unambiguously and in detail, and so discover a new evolutionary channel for the production of binary pulsating stars—these are inhabitants of the pulsational instability strip on the Hertzsprung–Russell diagram that mimic classical RR Lyrae variables, but have a completely different origin. These low mass pulsating stars could in principle increase the observed spread in luminosity of the RR Lyrae stars, hence affecting distance measurements based on them. Our calculations show that among 1,000 RR Lyrae stars one should expect just 2 such stars, so in practice they should not affect distance determinations to galaxies if these are made with relatively large samples of RR Lyrae stars.

Received 19 October 2011; accepted 10 February 2012.

- Mateo, M. Dwarf galaxies of the Local Group. *Annu. Rev. Astron. Astrophys.* **36**, 435–506 (1998).
- Bono, G. & Cignoni, M. in *Proceedings of the Symposium “The Three-Dimensional Universe with Gaia”* (eds Turon, C., O’Flaherty, K. S. & Perryman, M. A. C.) 659–666 (ESA SP-576, 2005).
- Szewczyk, O. *et al.* The Araucaria Project. The distance of the Large Magellanic Cloud from near-infrared photometry of RR Lyrae variables. *Astron. J.* **136**, 272–279 (2008).
- Soszyński, I. *et al.* The Optical Gravitational Lensing Experiment. The OGLE-III catalog of variable stars. XI. RR Lyrae stars in the Galactic Bulge. *Acta Astron.* **61**, 1–23 (2011).
- Christy, R. F. A study of pulsation in RR Lyrae models. *Astron. J.* **69**, 536–537 (1964).
- Smith, H. A. *RR Lyrae Stars* (Cambridge Univ. Press, 2004).
- Bono, G. *et al.* A pulsational approach to near-infrared and visual magnitudes of RR Lyr stars. *Mon. Not. R. Astron. Soc.* **344**, 1097–1106 (2003).
- Stepień, K. Evolution of cool close binaries — approach to contact. *Acta Astron.* **61**, 139–159 (2011).
- Stepień, K. Evolutionary status of late-type contact binaries. *Acta Astron.* **56**, 199–218 (2006).
- Kunder, A. *et al.* Period change similarities among the RR Lyrae variables in Oosterhoff I and Oosterhoff II globular systems. *Astron. J.* **141**, 15–28 (2011).
- Wilson, R. E. & Devinney, E. J. Realization of accurate close-binary light curves: application to MR Cygni. *Astrophys. J.* **166**, 605–620 (1971).
- Van Hamme, W. & Wilson, R. E. Third-body parameters from whole light and velocity curves. *Astrophys. J.* **661**, 1129–1151 (2007).
- Tognelli, E., Prada Moroni, P. G. & Deg’Innocenti, S. The Pisa pre-main sequence tracks and isochrones. A database covering a wide range of Z, Y, mass, and age values. *Astron. Astrophys.* **533**, A109 (2011).
- Asplund, M., Grevesse, N., Sauval, A. J. & Scott, P. The chemical composition of the Sun. *Annu. Rev. Astron. Astrophys.* **47**, 481–522 (2009).
- Prada Moroni, P. G. & Staniero, O. Very low-mass white dwarfs with a C-O core. *Astron. Astrophys.* **507**, 1575–1583 (2009).
- Bono, G., Caputo, F., Cassisi, S., Icerpi, R. & Marconi, M. Metal-rich RR Lyrae variables. II. The pulsational scenario. *Astrophys. J.* **483**, 811–825 (1997).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We acknowledge financial support from the Chilean Center for Astrophysics FONDAP, the BASAL Centro de Astrofísica y Tecnologías Afines (CATA), NSF, the Polish Ministry of Science (Ideas Plus), the Foundation for Polish Science (FOCUS, TEAM), and the GEMINI-CONICYT found. The OGLE project received funding from the European Research Council ‘Advanced Grant’ Program. We thank the staff astronomers at Las Campanas and ESO Paranal (program 287.D-5022(A)) who provided support in data acquisition.

**Author Contributions** G.P., photometric and spectroscopic observations and reductions, and data analysis. I.B.T., spectroscopic observations and reductions, radial velocity measurements, and data analysis. W.G., spectroscopic observations and data analysis. D.G., modelling and data analysis. K.S., theoretical models and analysis. G.B., theoretical models. A.U., photometric observations and reductions, and data analysis. I.S., photometric observations and reductions. B.P. and P.G.P.M., modelling. G.W.P., N.N., M.G., J.S. and P.K., analysis of the data. A.M. and I.U.R., spectroscopic observations. G.P. and W.G. worked jointly to draft the manuscript, with all authors reviewing and contributing to its final form.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to G.P. ([pietrzyn@astrouw.edu.pl](mailto:pietrzyn@astrouw.edu.pl)).



# A steady-state superradiant laser with less than one intracavity photon

Justin G. Bohnet<sup>1</sup>, Zilong Chen<sup>1</sup>, Joshua M. Weiner<sup>1</sup>, Dominic Meiser<sup>1†</sup>, Murray J. Holland<sup>1</sup> & James K. Thompson<sup>1</sup>

The spectral purity of an oscillator is central to many applications, such as detecting gravity waves<sup>1</sup>, defining the second<sup>2,3</sup>, ground-state cooling and quantum manipulation of nanomechanical objects<sup>4</sup>, and quantum computation<sup>5</sup>. Recent proposals<sup>6–9</sup> suggest that laser oscillators which use very narrow optical transitions in atoms can be orders of magnitude more spectrally pure than present lasers. Lasers of this high spectral purity are predicted to operate deep in the ‘bad-cavity’, or superradiant, regime, where the bare atomic linewidth is much less than the cavity linewidth. Here we demonstrate a Raman superradiant laser source in which spontaneous synchronization of more than one million rubidium-87 atomic dipoles is continuously sustained by less than 0.2 photons on average inside the optical cavity. By operating at low intracavity photon number, we demonstrate isolation of the collective atomic dipole from the environment by a factor of more than ten thousand, as characterized by cavity frequency pulling measurements. The emitted light has a frequency linewidth, measured relative to the Raman dressing laser, that is less than that of single-particle decoherence linewidths and more than ten thousand times less than the quantum linewidth limit typically applied to ‘good-cavity’ optical lasers<sup>10</sup>, for which the cavity linewidth is much less than the atomic linewidth. These results demonstrate several key predictions for future superradiant lasers, which could be used to improve the stability of passive atomic clocks<sup>3</sup> and which may lead to new searches for physics beyond the standard model<sup>11,12</sup>.

The most-coherent man-made oscillators are based on stable length references<sup>3,13</sup> that, for optical frequencies, exist as lasers stabilized to

optical reference cavities. Even in state-of-the-art stable lasers, the mirrors of the reference cavity to which they are stabilized vibrate as a result of thermal noise, causing time-integrated phase drifts that limit the laser linewidth to 0.125–0.3 Hz (refs 3, 14).

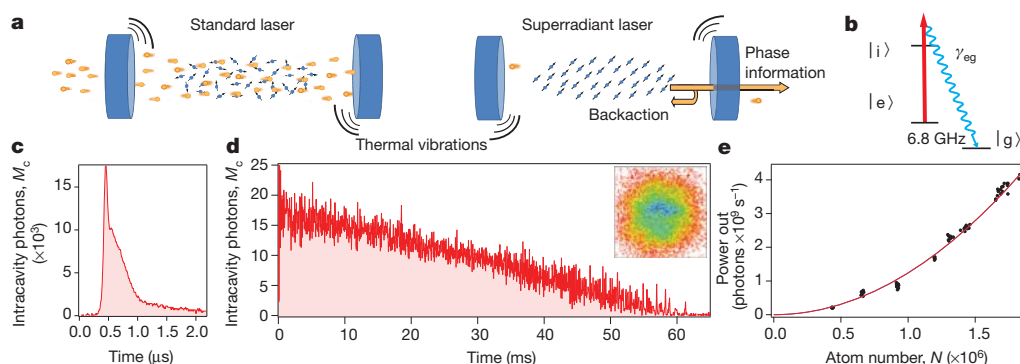
The fundamental linewidth of a laser is given by a generalized expression for the Schawlow–Townes full-width at half-maximum (FWHM),  $\Delta f_{\text{ST}}$ , in hertz<sup>15</sup> and which can be written in a simplified form as<sup>16</sup>

$$\Delta f_{\text{ST}} = \frac{1}{4\pi} \frac{hf}{P_{\text{out}}} \left( \frac{2\gamma_{\perp}\kappa}{2\gamma_{\perp} + \kappa} \right)^2 \quad (1)$$

Here  $P_{\text{out}}$  is the power exiting the cavity,  $f$  is the oscillation frequency,  $h$  is Planck’s constant and  $\kappa$  is the cavity power decay rate. The transverse decoherence rate of the lasing optical transition is  $\gamma_{\perp} = \gamma_{\text{eg}}/2 + 1/T_2$ , where  $\gamma_{\text{eg}}$  is the rate of decay from the excited state to the ground state and  $1/T_2$  parameterizes additional atomic dephasing mechanisms. In the good-cavity limit ( $2\gamma_{\perp} \gg \kappa$ ), the generalized linewidth expression reduces to  $\Delta f_{\text{GST}} = \kappa/4\pi M_c$ , where  $M_c$  is the average intracavity photon number (Supplementary Information).

If the cavity and atomic transition frequencies, respectively  $f_{\text{cav}}$  and  $f_{\text{atomic}}$ , are not identical, the system oscillates at a weighted average frequency  $f = (2\gamma_{\perp}f_{\text{cav}} + \kappa f_{\text{atomic}})/(2\gamma_{\perp} + \kappa)$ . The cavity frequency changes the oscillation frequency from the atomic transition frequency by an amount  $P \equiv df/df_{\text{cav}} = 2\gamma_{\perp}/(2\gamma_{\perp} + \kappa)$ , called the frequency pulling coefficient. In the good-cavity limit,  $P \approx 1$  (Fig. 1a).

In the bad-cavity, or superradiant, limit ( $2\gamma_{\perp} \ll \kappa$ ), the FWHM linewidth reduces to  $\Delta f_{\text{BST}} = \gamma_{\perp}^2/\pi\kappa M_c$ , a result intimately related to cavity narrowing using slowed light<sup>16,17</sup>. The collective atom–light excitation is stored predominantly in the gain medium, making the



**Figure 1 | A steady-state superradiant laser.** **a**, Left: in a standard, good-cavity laser far above threshold, many photons (yellow) circulate inside the cavity, extracting energy from the largely incoherent atomic gain medium (blue). Thermal vibrations of the mirror surfaces modulate the cavity resonance frequency, limiting the linewidth of the laser. Right: in a superradiant laser, the collective atomic dipole stores the coherence, and continuous stimulated emission can be achieved even with less than one photon in the cavity. The stimulation enables phase information to be extracted at a useful rate, and the small intracavity photon number leads to only weak cavity-induced backaction

on the collective atomic dipole. **b**, To mimic a narrow optical atomic transition, we dress the metastable ground state  $|e\rangle$  with a laser (red) to induce a spontaneous two-photon Raman transition to  $|g\rangle$  (blue), with tunable rate  $\gamma_{\text{eg}}$ . **c**, With no repumping light, a single superradiant pulse is emitted. **d**, With optical repumping back to  $|e\rangle$ , we observe quasi-continuous emission limited by atom loss. The atoms emit into a single spatial mode of the cavity ( $\text{TEM}_{00}$ ) imaged on a charge-coupled device (inset). **e**, The measured peak power output (black circles) scales as the square of the number of atoms. The red line is a quadratic fit to the data.

<sup>1</sup>JILA, NIST and University of Colorado, Department of Physics, University of Colorado, 440 UCB, Boulder, Colorado 80309, USA. <sup>†</sup>Present address: Tech-X Corporation, 5621 Arapahoe Avenue, Suite A, Boulder, Colorado 80303, USA.

atoms the primary carriers of phase information (Fig. 1a). The weak intracavity photon field acts mainly as a communication bus to drive spontaneous synchronization of the atomic dipoles and to extract information about the phase stored in the collective atomic dipole. The synchronized atomic dipoles radiate at an increased rate, a phenomenon known as superradiance or superfluorescence<sup>18</sup>. In addition, the frequency pulling coefficient becomes  $P \approx 2\gamma_{\perp}/\kappa \ll 1$ , drastically reducing the impact of noise in the cavity frequency. This isolation of the oscillator from the environment is the key to reducing the sensitivity of such a laser to thermal and technical noise. The goal of this work is to use a two-photon Raman transition as a model system for narrow optical transitions and to demonstrate operation deep into the optical bad-cavity regime.

At microwave frequencies, masers typically operate in the bad-cavity regime<sup>19</sup>. Because of the thermal photon occupation at microwave frequencies, cryogenic operation is required for optimal stability<sup>20</sup>. In the optical domain, the bad-cavity regime is less explored. HeNe and HeXe infrared gas lasers have been studied in the good-cavity/bad-cavity crossover regime  $2\gamma_{\perp}/\kappa = 6$  to  $0.2$ , with  $2\gamma_{\perp} \approx 2\pi \times 100$  to  $2\pi \times 500$  MHz, and with  $M_c \gg 1$  intracavity photons<sup>16,21</sup>. Microcavity lasers, such as vertical-cavity surface-emitting lasers, can operate near the bad-cavity regime<sup>22</sup>, but their usefulness as a state-of-the-art phase reference is limited by the linewidth broadening inherent in semiconductor gain media.

Raman lasers using laser-cooled, untrapped atoms have been operated in the good-cavity limit<sup>23</sup>, or in the bad-cavity limit without studying the spectral properties of the light, with  $M_c \gg 1$  intracavity photons, and with no distinction between the Raman dressing laser and repumping laser<sup>24–26</sup>. Additionally, optomechanical forces can make interpretation of the lasing mechanism difficult<sup>26</sup>.

Our experimental system, described in Supplementary Information and ref. 27, operates deep in the optical bad-cavity regime, with  $2\gamma_{\perp}/\kappa = 2 \times 10^{-5}$  to  $1 \times 10^{-3}$ . The effective excited-state scattering rate,  $\gamma_{eg} \approx 2$  to  $60 \text{ s}^{-1}$ , is comparable to those of some candidate optical clock candidates, and  $M_c = 0.2$  to  $60$ . The system consists of an ensemble of  $N \approx 10^6$   $^{87}\text{Rb}$  atoms confined to the TEM<sub>00</sub> mode of a medium-finesse ( $F = 710$ ) optical cavity with cavity power decay rate  $\kappa = 2\pi \times 11$  MHz. Optomechanical effects are suppressed by tightly trapping the atoms using a one-dimensional intracavity optical lattice with a wavelength of 823 nm. We observe no lasing at the vibrational sideband frequencies.

The cavity is coupled to an optically dressed, atomic ground state engineered to imitate a long-lived, optically excited atomic state<sup>6,7</sup> (Fig. 1b). The effective excited state,  $|e\rangle$ , and ground state,  $|g\rangle$ , are the magnetic-field-insensitive hyperfine clock ground states. The upper state,  $|e\rangle$ , is metastable and separated from  $|g\rangle$  by only 6.8 GHz, but we induce an optical Raman decay from  $|e\rangle$  to  $|g\rangle$  by applying a linearly polarized 795-nm dressing laser tuned 1 to 2 GHz higher in frequency than the optically excited intermediate state,  $|i\rangle$ . The cavity is tuned to resonance with the spontaneously emitted photon. The combination of detuning and dressing laser power sets a tunable single-particle decay rate,  $\gamma_{eg}$ , from  $|e\rangle$  to  $|g\rangle$ . The cycle is completed using 780-nm light to repump atoms incoherently from  $|g\rangle$  to  $|e\rangle$  at a rate proportional to the single-particle rate,  $w$ , of scattering from  $|g\rangle$ .

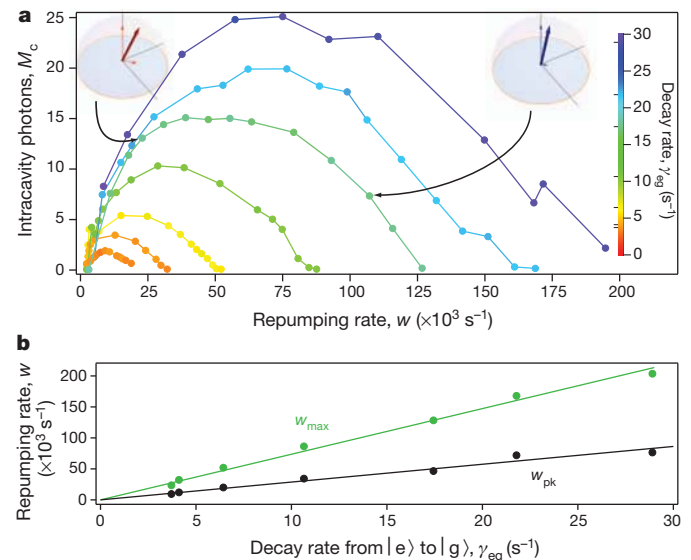
The superradiant linewidth,  $\Delta f_{\text{BST}}$ , reduces to  $C\gamma_{eg}/\pi$  at a peak repumping rate of  $w_{\text{pk}} = NC\gamma_{eg}/2$ . Here  $C = (2g_2)^2/\kappa\gamma_{eg}$  is the cooperativity and  $g_2$  is the single-particle, two-photon coupling of the atoms to the cavity emission mode after adiabatic elimination of the intermediate state  $|i\rangle$ . It should be possible to obtain  $\Delta f_{\text{BST}} \ll \gamma_{eg}$  by operating in the single-particle, weak-coupling regime,  $C \ll 1$ , a limit that contrasts with the trend in quantum optics to engineer systems with single-particle strong-coupling ( $C \gg 1$ ). It is predicted that strong collective coupling,  $NC \gg 1$ , is sufficient for self-sustained stimulated emission. In this work, we operate with  $C = 7.7(4) \times 10^{-3}$  and  $NC \approx 10^4$  (all reported uncertainties describe 68% confidence intervals).

The superradiant regime assumes continuous light generation, but optical superradiance is ordinarily regarded as an intrinsically pulsed process with no phase coherence between the pulses. A single superradiant pulse of photons is emitted if all atoms are initialized in  $|e\rangle$  and the dressing laser is applied (Fig. 1c). This pulse contains approximately 0.9(1) photons per atom. Without a pump to recycle the atoms to  $|e\rangle$ , this pulse is the full extent of the superradiance.

If instead we apply continuous optical repumping from  $|g\rangle$  to  $|e\rangle$  along with the dressing laser, we observe a transition to quasi-continuous superradiance lasting 20 to 140 ms. The integrated number of photons emitted into the cavity mode is approximately 35(4) photons per atom (Fig. 1d). The light source shuts off as atoms are lost from the lattice, most probably as a result of light-assisted atomic collisions (Supplementary Information) that can be suppressed using higher-dimensional lattices. We are able to access the continuous optical superradiant regime by engineering a system both with a large imbalance in the ratio of single-particle decay,  $\gamma_{eg}$ , to repumping rate,  $w$ , and with long coherence times  $T_2$  due to laser cooling and trapping.

Ordinary single-particle fluorescence from an atomic ensemble is too weak to be useful as a phase reference, but the rate of emission into a single spatial mode from the collective dipole is increased by approximately  $N$ . The predicted peak flux of photons from the cavity is  $\dot{M}_{\text{pk}} = RN^2C\gamma_{eg}/8$ , where  $R = 1$  for the two-level model of ref. 8 and a dot denotes a time derivative. Typically,  $R \approx 0.2$  for our multilevel system owing to the necessity of optically repumping the atoms through several intermediate Zeeman states. We find that the peak output power scales as  $N^2$  (Fig. 1e). For this data,  $2\gamma_{\perp}/\kappa \approx w/\kappa = 3.9(8) \times 10^{-4}$  to  $1.4(3) \times 10^{-3}$ .

Repumping non-collectively scatters photons transverse to the cavity mode, causing collapse of individual atomic wavefunctions, whose pseudo-spins are then effectively torqued to rejoin the collective dipole by the weak intracavity field. If the repumping rate is greater than  $w_{\text{max}} = NC\gamma_{eg}$ , the torquing caused by the cavity field cannot replenish the collective dipole quickly enough relative to the rate of repumping induced collapse<sup>6</sup>. The predicted peak output power occurs for  $w_{\text{pk}} = w_{\text{max}}/2$ . The quenching and linear scaling of  $w_{\text{max}}$  and  $w_{\text{pk}}$  with  $\gamma_{eg}$  are shown in Fig. 2. The fitted and predicted slopes agree to within the 20% uncertainty in the predicted values.



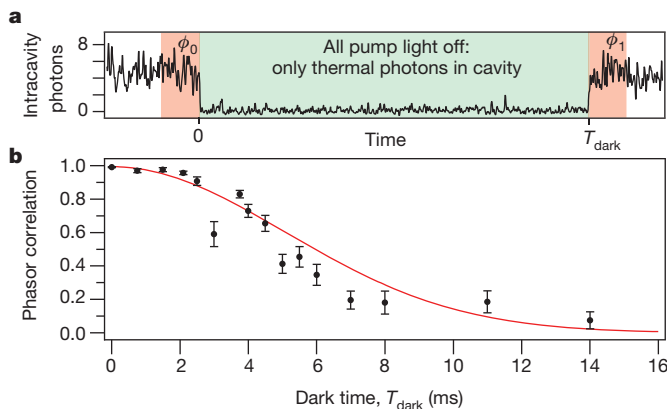
**Figure 2 | Repumping-induced quenching.** **a**, Average number of intracavity photons,  $M_c$ , versus the repumping rate,  $w$ , for different decay rates,  $\gamma_{eg}$ . The inferred Bloch vector is shown for two operating points. **b**, The repumping rate above which superradiance is quenched,  $w_{\text{max}}$  (green), and the repumping rate at peak output power,  $w_{\text{pk}}$  (black), scale linearly with  $\gamma_{eg}$ , as shown by linear fits to the data.

To show explicitly that the phase memory of the system resides in the collective dipole, we measure the phase of the emitted light just before,  $\phi_0$ , and after,  $\phi_1$ , a dark period of duration  $T_{\text{dark}}$ , in which both the dressing and repumping lasers are shut off (Fig. 3). During this period, any photons in the cavity exit with exponential time constant  $1/\kappa = 14$  ns, rapidly erasing all phase information stored in the light mode. Whereas the phase  $\phi_0$  is random from one trial to the next, the phase difference within a single trial,  $\phi_1 - \phi_0$ , is highly correlated between trials, demonstrating that phase information is being stored within the collective atomic dipole.

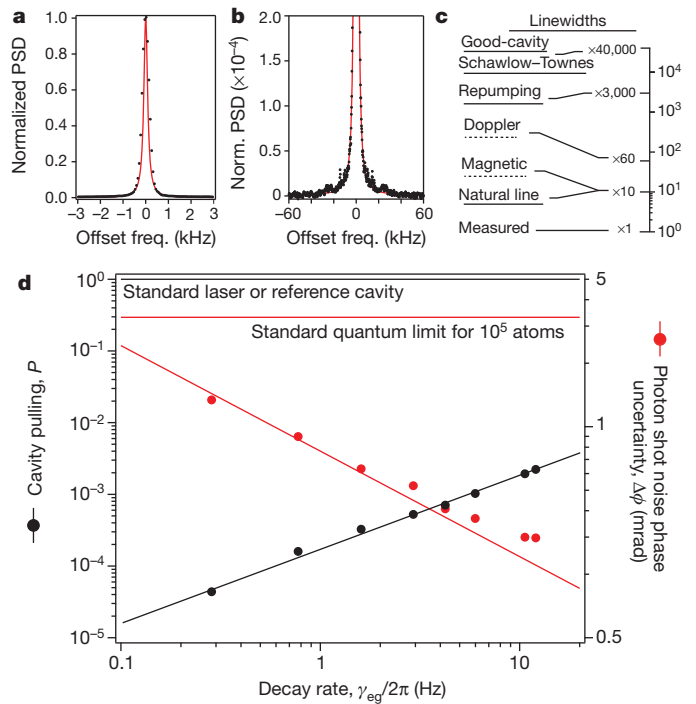
In Fig. 4a, we present the power spectrum (PSD) of the emitted light heterodyned with the dressing laser with  $M_c = 0.20(2)$  photons in the cavity mode. The measured Gaussian FWHM, 350(25) Hz, is four orders of magnitude smaller than the good-cavity Schawlow–Townes linewidth limit,  $f_{\text{GST}}$ , assuming one intracavity photon. The measured linewidth is also less than single-particle decoherence rates arising from the repumping-broadened linewidth,  $w$ , inhomogeneous magnetic field broadening and Doppler decoherence (Fig. 4c). Such broadening and decoherence measurements are performed in a secondary configuration described in Supplementary Information. We believe that the predicted superradiant linewidth, 2(1) mHz, is not observed primarily as a result of dispersive tuning of the cavity mode frequency caused by changes in atom number between individual experimental trials, an effect that is present in this Raman system but not in proposed active optical clocks.

To probe fundamental noise limits further, we note that quantum phase diffusion, as well as single-particle total excited state decay (rate,  $\gamma_e$ ) and repumping broadening (rate,  $w$ ), all produce a Lorentzian contribution to the PSD. After low-frequency spectral components have been removed as described in Supplementary Information, a Lorentzian fit that excludes offset frequencies of less than 4 kHz yields a FWHM of 4.5(5) Hz (Fig. 4b). This is still larger than the predicted linewidth, 10(2) mHz, for this particular data set. We believe that the linewidth is again limited by atomic population noise as described above. However, the measured linewidth is less than the total Raman decay linewidth,  $\gamma_e/2\pi = 48(16)$  Hz (and the repumping linewidth,  $w/2\pi = 14(2)$  kHz), demonstrating that a collective oscillation can have lower phase noise than can be obtained from a collection of independent atoms fluorescing in free space, as predicted for  $C < 1$ .

In Fig. 4d, we demonstrate cavity pulling coefficients  $P = 4 \times 10^{-5}$  to  $2 \times 10^{-3}$  for a range of  $\gamma_{\text{eg}}$  values comparable to the linewidth of several atoms suitable for use as optical clocks. The pulling is measured



**Figure 3 | Phase coherence maintained with no intracavity photons.** **a**, The dressing and repumping lasers are shut off for some variable dark time,  $T_{\text{dark}}$ . The phase of the light relative to the dressing laser is measured shortly before and after the dark time. **b**, Phase correlation function, defined in Supplementary Information. For  $T_{\text{dark}} < 7$  ms, the phases are correlated because the collective atomic dipole seeds re-establishment of superradiant emission. At longer dark times, the collective atomic dipole dephases and superradiant emission restarts with a random phase. A Gaussian fit is shown in red. Error bars, 1 s.d. calculated from  $n \approx 75$  trials.



**Figure 4 | Beyond good-cavity, optical laser stability.** **a**, Average PSD of the heterodyne signal of the emitted light (black circles) with  $M_c = 0.20(2)$  intracavity photons, and a Gaussian fit (red line). **b**, Average PSD of the heterodyne signal of the emitted light (black circles) with  $M_c = 31(5)$ , and a Lorentzian fit (red line) that excludes offset frequencies of less than 4 kHz. **c**, Comparison of various characteristic linewidths. Gaussian noise processes are compared to the measured Gaussian FWHM (dashed), and Lorentzian noise processes are compared to the measured Lorentzian FWHM (solid). **d**, As the decay rate  $\gamma_{\text{eg}}$  decreases, the atomic dipole becomes more isolated from the mirrors, as shown by directly measuring the cavity frequency pulling coefficient  $P$  (black circles). Because of the collectively enhanced emission rate, phase information is still coupled out of the system at a sufficient rate that the photon shot noise phase uncertainty limit (in a 1 kHz bandwidth, red circles) would be sufficient for spectroscopy below the standard quantum limit for  $10^5$  atoms. The pulling and phase noise data are shown with fits scaling as  $\gamma_{\text{eg}}$  and  $1/\sqrt{\gamma_{\text{eg}}}$ , respectively.

by applying small changes,  $df_{\text{cav}}$ , in the cavity resonance frequency and measuring the frequency shift,  $df$ , of the light exiting the cavity. The measured pulling coefficients agree reasonably well with the predicted pulling coefficient,  $P_{\text{pred}} = 2\gamma_{\perp}/\kappa$ ;  $P/P_{\text{pred}} = 1.44(22)$ . Assuming the suppression of cavity pulling measured in this work, current optical cavities stabilized to  $\sim 1$  Hz are sufficient to reach a linewidth of 1 mHz with the appropriate atom.

To decouple the collective atomic dipole from the environment, we can choose an atom with a small  $\gamma_{\text{eg}}$  value. However, the emitted light power would also decrease, making the light source too weak to phase stabilize even state-of-the-art narrow lasers. For a 1-kHz bandwidth, we observe a flux of photons sufficient to obtain a quantum-limited phase uncertainty of  $\Delta\phi \approx 1$  mrad or less (Fig. 4d and Supplementary Information). This phase uncertainty and servo bandwidth would be sufficient to stabilize a bright, 1-Hz-linewidth laser down to 1 mHz, making it possible to do spectroscopy below the standard quantum limit for ensembles of  $10^5$  atoms in a state-of-the-art optical lattice clock.

This work explores a novel regime of optical laser physics in which a highly coherent collective atomic dipole replaces the optical cavity as the high-quality-factor oscillator. Photons are essential in that they establish atomic correlations, but a macroscopic number of photons need not occupy the intracavity mode. This light source, with  $M_c < 1$ ,  $N \gg 1$  and  $C \ll 1$ , operates in the opposite limit of the single-atom laser<sup>28</sup>, for which  $N = 1$ ,  $M_c > 1$  and  $C \gg 1$ . Our results are particularly



relevant for precision measurement science, because we demonstrate such a source operating in the ‘no-photon’ limit,  $M_c < 1$ , which may be necessary for superradiant lasers with linewidths of  $< 1$  mHz. In addition, we show that because our device is an active optical oscillator, stability in excess of the limits imposed by single-particle decoherence rates can be achieved. Additional experiments are needed to study photon statistics, atom loss and possible unforeseen noise that our experiment was unable to resolve at low offset frequencies.

A future active optical frequency source based on the physical principles demonstrated here might have wide-ranging impact. A 1-mHz source would immediately improve the stability of atomic clocks, because a 1-mHz linewidth may prove sufficient to exhaust the present coherence times offered by passive optical lattice clocks<sup>2</sup>, with implications for the Global Positioning System, communications and precision measurements. Such a robust frequency standard might be deployed on portable platforms for geodetic surveys<sup>29</sup>, or its long coherence length, which is of the order of the Earth–Sun separation, might be applied to space-based optical interferometry.

Received 2 September 2011; accepted 2 February 2012.

- Cagnoli, G. *et al.* Very high  $Q$  measurements on a fused silica monolithic pendulum for use in enhanced gravity wave detectors. *Phys. Rev. Lett.* **85**, 2442–2445 (2000).
- Ludlow, A. D. *et al.* Sr lattice clock at  $1 \times 10^{-16}$  fractional uncertainty by remote optical evaluation with a Ca clock. *Science* **319**, 1805–1808 (2008).
- Jiang, Y. Y. *et al.* Making optical atomic clocks more stable with  $10^{-16}$ -level laser stabilization. *Nature Photon.* **5**, 158–161 (2011).
- Teufel, J. D. *et al.* Sideband cooling of micromechanical motion to the quantum ground state. *Nature* **475**, 359–363 (2011).
- Leibfried, D., Blatt, R., Monroe, C. & Wineland, D. Quantum dynamics of single trapped ions. *Rev. Mod. Phys.* **75**, 281–324 (2003).
- Meiser, D., Ye, J., Carlson, D. R. & Holland, M. J. Prospects for a millihertz-linewidth laser. *Phys. Rev. Lett.* **102**, 163601–163604 (2009).
- Chen, J. Active optical clock. *Chin. Sci. Bull.* **54**, 348–352 (2009).
- Meiser, D. & Holland, M. J. Steady-state superradiance with alkaline-earth-metal atoms. *Phys. Rev. A* **81**, 033847–033850 (2010).
- Meiser, D. & Holland, M. J. Intensity fluctuations in steady-state superradiance. *Phys. Rev. A* **81**, 063827–063833 (2010).
- Schawlow, A. L. & Townes, C. H. Infrared and optical masers. *Phys. Rev.* **112**, 1940–1949 (1958).
- Blatt, S. *et al.* New limits on coupling of fundamental constants to gravity using  $^{87}\text{Sr}$  optical lattice clocks. *Phys. Rev. Lett.* **100**, 140801–140804 (2008).
- Fortier, T. M. *et al.* Precision atomic spectroscopy for improved limits on variation of the fine structure constant and local position invariance. *Phys. Rev. Lett.* **98**, 070801–070804 (2007).
- Hartnett, J. & Nand, N. Ultra-low vibration pulse-tube cryocooler stabilized cryogenic sapphire oscillator with  $10^{-16}$  fractional frequency stability. *IEEE Trans. Microw. Theory Tech.* **58**, 3580–3586 (2010).
- Young, B. C., Cruz, F. C., Itano, W. M. & Bergquist, J. C. Visible lasers with subhertz linewidths. *Phys. Rev. Lett.* **82**, 3799–3802 (1999).
- Kolobov, M. I., Davidovich, L., Giacobino, E. & Fabre, C. Role of pumping statistics and dynamics of atomic polarization in quantum fluctuations of laser sources. *Phys. Rev. A* **47**, 1431–1446 (1993).
- Kuppens, S. J. M., van Exter, M. P. & Woerdman, J. P. Quantum-limited linewidth of a bad-cavity laser. *Phys. Rev. Lett.* **72**, 3815–3818 (1994).
- Wang, H., Goorskey, D. J., Burkett, W. H. & Xiao, M. Cavity-linewidth narrowing by means of electromagnetically induced transparency. *Opt. Lett.* **25**, 1732–1734 (2000).
- Dicke, R. H. Coherence in spontaneous radiation processes. *Phys. Rev.* **93**, 99–110 (1954).
- Kleppner, D., Goldenberg, H. M. & Ramsey, N. F. Theory of the hydrogen maser. *Phys. Rev.* **126**, 603–615 (1962).
- Benmessai, K. *et al.* Measurement of the fundamental thermal noise limit in a cryogenic sapphire frequency standard using bimodal maser oscillations. *Phys. Rev. Lett.* **100**, 233901–233904 (2008).
- Kuppens, S., van Exter, M., Woerdman, J. & Kolobov, M. Observation of the effect of spectrally inhomogeneous gain on the quantum-limited laser linewidth. *Opt. Commun.* **126**, 79–84 (1996).
- Björk, G., Heitmann, H. & Yamamoto, Y. Spontaneous-emission coupling factor and mode characteristics of planar dielectric microcavity lasers. *Phys. Rev. A* **47**, 4451–4463 (1993).
- Vrijsen, G., Hosten, O., Lee, J., Bernon, S. & Kasevich, M. A. Raman lasing with a cold atom gain medium in a high-finesse optical cavity. *Phys. Rev. Lett.* **107**, 063904–063907 (2011).
- Hilico, L., Fabre, C. & Giacobino, E. Operation of a cold-atom laser in a magneto-optical trap. *Europhys. Lett.* **18**, 685–688 (1992).
- Guerin, W., Michaud, F. & Kaiser, R. Mechanisms for lasing with cold atoms as the gain medium. *Phys. Rev. Lett.* **101**, 093002–093005 (2008).
- Chan, H. W., Black, A. T. & Vuletić, V. Observation of collective-emission-induced cooling of atoms in an optical cavity. *Phys. Rev. Lett.* **90**, 063003–063006 (2003).
- Chen, Z., Bohnet, J. G., Sankar, S. R., Dai, J. & Thompson, J. K. Conditional spin squeezing of a large ensemble via the vacuum Rabi splitting. *Phys. Rev. Lett.* **106**, 133601–133604 (2011).
- McKeever, J., Boca, A., Boozer, A. D., Buck, J. R. & Kimble, H. J. Experimental realization of a one-atom laser in the regime of strong coupling. *Nature* **425**, 268–271 (2003).
- Leibrandt, D. R., Thorpe, M. J., Bergquist, J. C. & Rosenband, T. Field-test of a robust, portable, frequency-stable laser. *Opt. Express* **19**, 10278–10286 (2011).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank J. Ye and A. M. Rey for discussions. J.G.B., Z.C., J.M.W. and J.K.T. acknowledge support from NSF PFC, NIST and ARO. M.J.H. and D.M. acknowledge support from the DARPA QuASAR programme through a grant from ARO and from NSF. J.G.B. acknowledges support from NSF GRF, and Z.C. acknowledges support from A\*STAR Singapore.

**Author Contributions** J.G.B., Z.C., J.M.W. and J.K.T. designed and built the experiment. J.G.B. and Z.C. performed the measurements. J.G.B., Z.C., J.M.W. and J.K.T. analysed the results. D.M., M.J.H. and J.K.T. provided the theoretical analysis. J.G.B. and J.K.T. wrote the manuscript. All authors discussed the results and text of the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to J.K.T. (jkt@jila.colorado.edu).

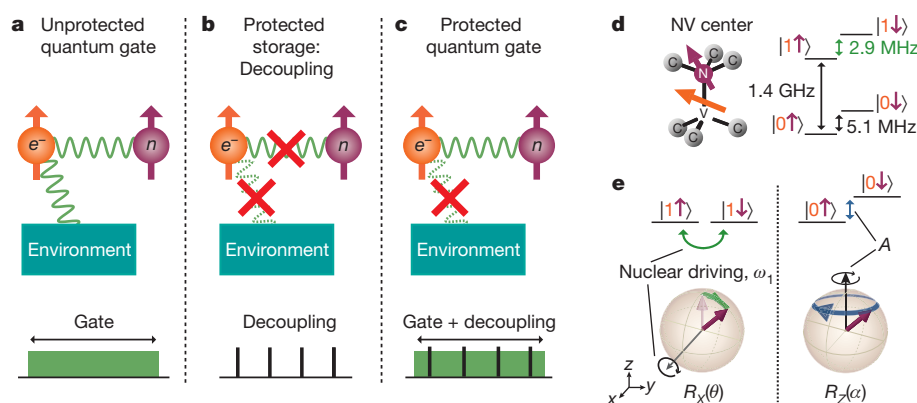
# Decoherence-protected quantum gates for a hybrid solid-state spin register

T. van der Sar<sup>1</sup>, Z. H. Wang<sup>2</sup>, M. S. Blok<sup>1</sup>, H. Bernien<sup>1</sup>, T. H. Taminiau<sup>1</sup>, D. M. Toyli<sup>3</sup>, D. A. Lidar<sup>4</sup>, D. D. Awschalom<sup>3</sup>, R. Hanson<sup>1</sup> & V. V. Dobrovitski<sup>2</sup>

Protecting the dynamics of coupled quantum systems from decoherence by the environment is a key challenge for solid-state quantum information processing<sup>1,2</sup>. An idle quantum bit (qubit) can be efficiently insulated from the outside world by dynamical decoupling<sup>3</sup>, as has recently been demonstrated for individual solid-state qubits<sup>4–9</sup>. However, protecting qubit coherence during a multi-qubit gate is a non-trivial problem<sup>3,10,11</sup>: in general, the decoupling disrupts the interqubit dynamics and hence conflicts with gate operation. This problem is particularly salient for hybrid systems<sup>12–22</sup>, in which different types of qubit evolve and decohere at very different rates. Here we present the integration of dynamical decoupling into quantum gates for a standard hybrid system, the electron–nuclear spin register. Our design harnesses the internal resonance in the coupled-spin system to resolve the conflict between gate operation and decoupling. We experimentally demonstrate these gates using a two-qubit register in diamond operating at room temperature. Quantum tomography reveals that the qubits involved in the gate operation are protected as accurately as idle qubits. We also perform Grover’s quantum search algorithm<sup>1</sup>, and achieve

fidelities of more than 90% even though the algorithm run-time exceeds the electron spin dephasing time by two orders of magnitude. Our results directly allow decoherence-protected interface gates between different types of solid-state qubit. Ultimately, quantum gates with integrated decoupling may reach the accuracy threshold for fault-tolerant quantum information processing with solid-state devices<sup>1,11</sup>.

Decoherence is a major hurdle in realizing scalable quantum technologies in the solid state. The interqubit dynamics that implement the quantum logic are unavoidably affected by uncontrolled couplings to the solid-state environment, preventing high-fidelity gate performance (Fig. 1a). Dynamical decoupling<sup>3</sup>, a technique that uses fast qubit flips to average out the interactions with the environment, is a powerful and practical tool for mitigating decoherence<sup>4–11,23–25</sup>. This approach is particularly promising for the emerging class of hybrid quantum architectures<sup>12–22</sup>, in which different types of qubit, such as electron and nuclear spins, superconducting resonators and nanomechanical oscillators, perform different functions. Dynamical decoupling allows each qubit type to be decoupled at its own rate, ensuring uniform coherence protection.



**Figure 1 | Quantum gate operation in the presence of decoherence.**

**a–c**, Challenge of high-fidelity quantum gates for qubits (orange, electron spin; purple, nuclear spin) coupled to a decohering environment. **a**, Without decoherence protection, the fidelity of two-qubit gates is limited by interactions with the environment. **b**, Dynamical decoupling efficiently preserves the qubit coherence (protected storage) by turning off the interaction between the qubit and its environment. However, this generally also decouples the qubit from other qubits and prevents two-qubit gate operations. If the decoupling and the gate are separated in time, the unprotected gate is still susceptible to decoherence-induced errors. **c**, The goal is to perform dynamical decoupling during the gate operation, thus ensuring that the gates are protected against decoherence. The gate operation should therefore be compatible with decoupling. The dephasing rate of the nuclear spin is negligible in our experiments. However, nuclear spin protection can easily be incorporated using another layer of decoupling. **d**, The two-qubit system used in this work: a nitrogen–vacancy (NV) colour centre in diamond carries an electron spin  $S = 1$

(orange) coupled to a  $^{14}\text{N}$  nuclear spin  $I = 1$  (purple). The states of the electronic qubit,  $|0\rangle$  and  $|1\rangle$ , are split by 1.4 GHz in an external field  $B_0 = 510$  G. The states  $|0\rangle$  and  $|1\rangle$  are split by 5.1 MHz owing to nuclear quadrupole and Zeeman interactions. The hyperfine coupling yields an additional splitting, such that the levels  $|1\rangle$  and  $|1\rangle$  are separated by 2.9 MHz. The Rabi driving is applied in resonance with this transition. **e**, Dynamics of the electron–nuclear spin system in the limit  $\omega_1 \ll A$ , visualized in a coordinate frame that rotates with frequency 1.4 GHz in the electron spin subspace and frequency 2.9 MHz in the nuclear spin subspace. In this frame, the states  $|1\rangle$  and  $|1\rangle$  have the same energy. The Rabi driving field, which is directed along the  $x$  axis, coherently rotates the nuclear spin if the electronic qubit is in  $|1\rangle$  (the resulting rotation around the  $x$  axis by angle  $\theta$  is denoted  $R_x(\theta)$ ). However, the Rabi driving is negligible for the states  $|0\rangle$  and  $|0\rangle$ , which differ in energy by  $A = 2\pi \times 2.16$  MHz. The phase accumulation between  $|0\rangle$  and  $|0\rangle$  corresponds to a coherent rotation of the nuclear spin around the  $z$  axis with frequency  $A$  (denoted  $R_z(\alpha)$ , where  $\alpha$  is the rotation angle).

<sup>1</sup>Kavli Institute of Nanoscience, Delft University of Technology, PO Box 5046, 2600 GA Delft, The Netherlands. <sup>2</sup>Ames Laboratory and Iowa State University, Ames, Iowa 50011, USA. <sup>3</sup>Center for Spintronics and Quantum Computation, University of California, Santa Barbara, California 93106, USA. <sup>4</sup>Departments of Electrical Engineering, Chemistry, and Physics, and Center for Quantum Information Science and Technology, University of Southern California, Los Angeles, California 90089, USA.

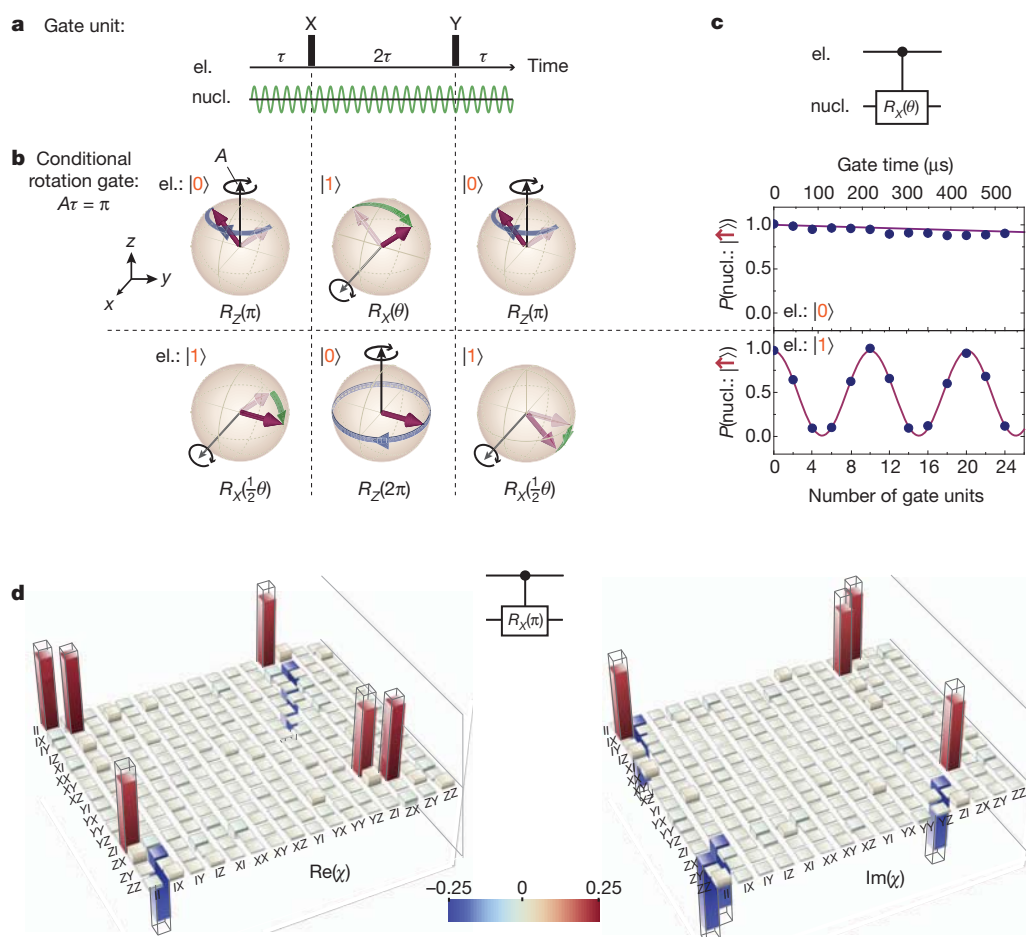
However, combining dynamical decoupling with quantum gate operations is generally problematic, because decoupling does not distinguish the desired interqubit interaction from the coupling to the decohering environment, and in general cancels both (Fig. 1b). For hybrid systems, where coherence and control timescales greatly differ among the different qubit types, this problem cannot be remedied by encoding<sup>10</sup> or by the synchronized application of decoupling pulses<sup>3,11</sup> as these require all qubits to be controlled on similar timescales. As a result, decoherence-protected quantum gates in hybrid systems have thus far remained elusive.

Here we present a design that allows the integration of decoupling into gate operation for hybrid quantum architectures. We demonstrate such decoherence-protected gates in a prototype hybrid quantum system: a two-qubit register consisting of an electron and a nuclear spin (Fig. 1c). The key idea is to adapt the time intervals between the electron decoupling pulses precisely to the nuclear spin dynamics. When combined with continuous nuclear spin driving, this yields highly selective rotations of the nuclear spin, and at the same time the electron spin is dynamically protected as explained below. This design preserves all of the advantages of dynamical decoupling without requiring additional qubits or controllable interqubit couplings. It can be readily implemented to yield decoherence-protected quantum gates in a range of hybrid systems, such as various electron–nuclear spin

registers<sup>12–16,19</sup>, and interface gates between the qubits and a spin–chain quantum data bus<sup>21,22</sup>.

We experimentally demonstrate the scheme on a single nitrogen–vacancy centre in diamond<sup>12,15,26–28</sup>, where the two qubits are represented by the electron spin and the host <sup>14</sup>N nuclear spin (Fig. 1d and Supplementary Information). An entangling gate between these qubits can be implemented using the hyperfine interaction, described by the Hamiltonian  $\hat{H}_{\text{hf}} = A\hat{I}_Z\hat{S}_Z$ , where  $A = 2\pi \times 2.16$  MHz for nitrogen–vacancy centres<sup>27,28</sup> and  $\hat{I}_Z$  and  $\hat{S}_Z$  are respectively the nuclear and electron spin operators. For an unprotected gate, of duration  $T_G = \pi/A$ , the fidelity is limited by the electron spin dephasing, which is dominated either by the bath of <sup>13</sup>C nuclear spins<sup>12</sup> (as in the experiments here) or by the electron spins of substitutional nitrogen atoms<sup>26</sup>. Decoupling applied to the electron spin suppresses decoherence but also disrupts the electron–nucleus coupling. At the same time, synchronized application of the decoupling pulses to both the electronic and the nuclear spin qubit is problematic, because a nuclear spin rotation takes longer ( $>10\mu\text{s}$  in typical experiments<sup>16,27</sup>) than the electron spin's dephasing time,  $T_2^*$  (0.5–5  $\mu\text{s}$  for nitrogen–vacancy centres<sup>12,26</sup>; Supplementary Information).

In our gate design, Rabi driving is applied at the frequency of the nuclear spin transition corresponding to the electron spin state  $|1\rangle$  ( $m_S = -1$ ; Fig. 1d). In the rotating frame, the two spins are described



**Figure 2 | Decoherence-protected quantum gates for an electron–nuclear spin register.** **a**, Basic unit of the decoherence-protected gate, consisting of dynamical decoupling of the electron (el.) spin and continuous nuclear (nucl.) spin driving. X and Y respectively denote electron  $\pi$ -pulses around the x and y axes, and  $\tau$  is the delay time. **b**, Bloch spheres showing the nuclear spin dynamics, conditioned on the initial electronic spin state, during a single gate unit. **c**, Decoherence-protected controlled rotation of the nuclear spin. Plotted is the measured probability of the nuclear spin to be in state  $|\uparrow\rangle$  as a function of

the number of applied gate units, for the two different electron spin input states. Lines are fits taking into account longitudinal relaxation of the electron spin, which was measured independently. For details of the fitting and dynamics of other input states, see Supplementary Information. Error bars corresponding to 1 s.d. are smaller than the data points. **d**, Measured process matrix,  $\chi$ , of the decoherence-protected CNOT gate. The transparent bars indicate the values of the matrix elements for the ideal gate.



by the Hamiltonian  $\hat{H} = A\hat{I}_Z\hat{S}_Z + \omega_1\hat{I}_X$ , where  $\omega_1$  is the nuclear Rabi frequency. The nuclear spin dynamics is then conditioned on the electron spin state (Fig. 1e and Supplementary Information): the nuclear spin undergoes driven rotation around the  $x$  axis if the electron spin is in  $|1\rangle$  and precesses around the  $z$  axis if the electron is in  $|0\rangle$  ( $m_S = 0$ ). Although Fig. 1d shows the case with  $\omega_1 \ll A$ , which is relevant to our experiments, the scheme works for arbitrary values of  $\omega_1$  and  $A$  (Supplementary Information).

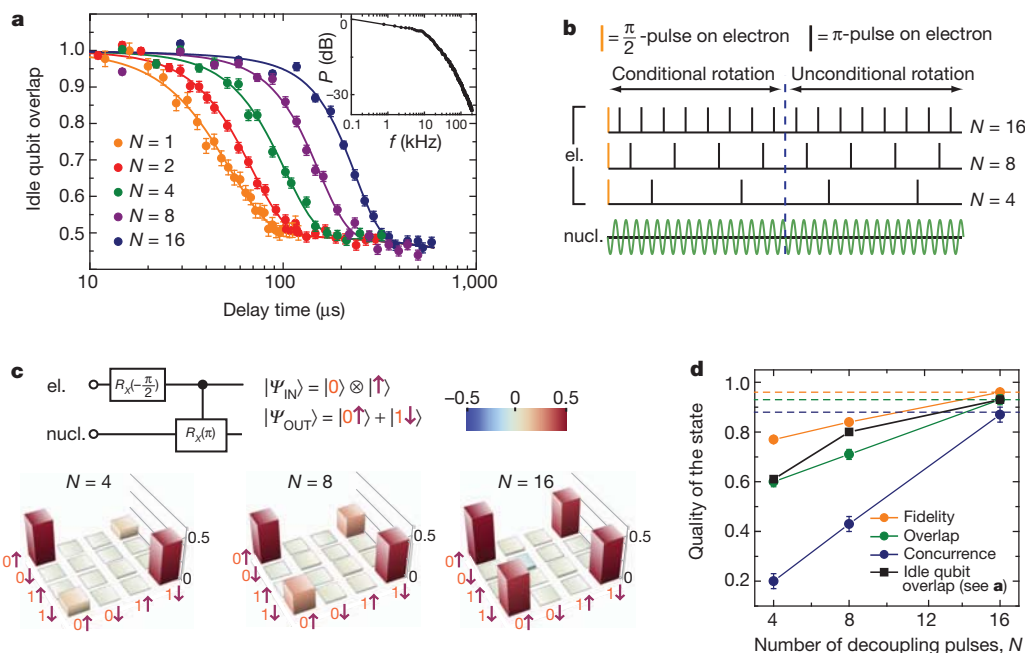
In parallel with driving the nuclear spin, we decouple the electron spin from the environment using short microwave pulses that constantly switch the electron spin between states  $|0\rangle$  and  $|1\rangle$  (refs 4, 5, 7). The full decoupling sequence consists of repeating the basic unit  $\tau$ -X- $2\tau$ -Y- $\tau$  (Fig. 2a), where X and Y are decoupling pulses that flip the electron spin around the  $x$  and  $y$  axes, respectively, and  $2\tau$  is the interpulse delay. To implement an entangling gate, the nuclear spin dynamics over the full gate unit must depend only on the initial state of the electron spin and cannot be disrupted by the electron's fast switching.

At first, such a condition seems to contradict the very design of the decoupling sequence. However, it may be fulfilled in a special situation where the interpulse delay matches the hyperfine coupling with high precision (at the nanosecond level in our experiments; see Supplementary Information), such that the decoupling pulses are applied in resonance with the rotation of the nuclear spin. This situation provides an opportunity for highly selective coupling of the fast qubit to the targeted slow qubit, while decoupling all others. Experimentally, it is achieved by setting  $\tau = (2n + 1)\pi/A$  with integer  $n$ , so that the nuclear spin rotates by  $180^\circ$  around the  $z$  axis during the interval  $\tau$ , if the electron is in state  $|0\rangle$  during this time. If the electron is in the state  $|0\rangle$  during the time interval  $2\tau$ , however, the nuclear spin undergoes a

$360^\circ$  rotation that has zero net effect. The overall dynamics during one full gate unit are shown in Fig. 2b: depending on the initial state of the electron, the nuclear spin rotates around the  $x$  axis by the angle  $\theta = 4\omega_1\tau$  in the clockwise or anticlockwise direction (conditional rotation). An unconditional rotation of the nuclear spin, independent of the electron spin state, is constructed from the same gate unit by choosing  $\tau = 2n\pi/A$  (Supplementary Fig. 3). From the conditional and unconditional rotations, we can construct a complete set of gates for the two-qubit register.

We implement the controlled-rotation gate using a conditional nuclear spin rotation followed by an unconditional rotation by the same angle. The experimental data in Fig. 2c confirm the selectivity of this gate: as the number of gate units increases, the nuclear spin rotates around the  $x$  axis if the electron spin is initially in state  $|1\rangle$ , and does nothing if the electron spin is initially in state  $|0\rangle$ . When the total nuclear rotation angle equals  $180^\circ$ , the gate corresponds to the controlled-NOT (CNOT) gate (neglecting a  $\pi/2$  phase shift of the electron spin, which is not relevant here and can be easily corrected). To characterize the CNOT gate fully, we performed quantum process tomography of its action (Fig. 2d). We find an overall process fidelity of  $F_p = \text{Tr}(\chi_{\text{ideal}}\chi) = 83(1)\%$ , where  $\chi$  is the measured process matrix,  $\chi_{\text{ideal}}$  is the process matrix of the ideal CNOT gate and the parenthetical error is the s.d. This number is a lower bound on the true process fidelity, as it includes errors from imperfect initialization (estimated to be about 5% for this experiment) and errors in the pulses used for state preparation and read-out (estimated to be a few per cent; see Supplementary Information).

A crucial step in this work is testing our design in the presence of stronger decoherence and confirming that it ensures efficient protection during the gate operation. To change the level of decoherence in a



**Figure 3 | Performance of the CNOT gate in the presence of strong decoherence.** **a**, Dynamical decoupling of the electron spin in the presence of artificially generated low-frequency noise, which is injected into the sample through the same microwave stripline used for the qubit control pulses. The coherence time of the electron spin is effectively prolonged by increasing the number of decoupling pulses,  $N$ , in the applied microwave pulse sequence  $\pi/2 - (\tau - \pi - \tau)^N - \pi/2$ , where  $\tau$  is the delay time. The points are experimental data and the lines are fits. Fit details are given in Supplementary Information. Inset, power spectrum of the injected noise ( $P$ , power;  $f$ , frequency; log scale). **b**, Pulse sequences used to test the performance of the CNOT gate for an increasing number of decoupling pulses,  $N$ , and a correspondingly decreasing interpulse

delay,  $\tau$  (total gate time is fixed at 120  $\mu$ s). **c**, Density matrices (real part) of the states created with the pulse sequences shown in **b**. A rotation of the electron spin by  $-\pi/2$  about the  $x$  axis prepares the state  $(|0\rangle + i|1\rangle) \otimes |\uparrow\rangle$ , which is transformed into the Bell state  $|0\rangle\uparrow + |1\rangle\downarrow$  by a 120- $\mu$ s CNOT gate. **d**, Fidelity, overlap and concurrence of the states shown in **c**, as functions of the number of decoupling pulses in the CNOT gate. The points are measured data connected by solid lines. The dashed horizontal lines correspond to the performance of the same gate with 16 decoupling pulses in the absence of the artificially generated noise. For comparison, the state overlap for decoupling of an idle electron spin, not involved in gate operation, is also shown (black squares and line). All errors, 1 s.d.

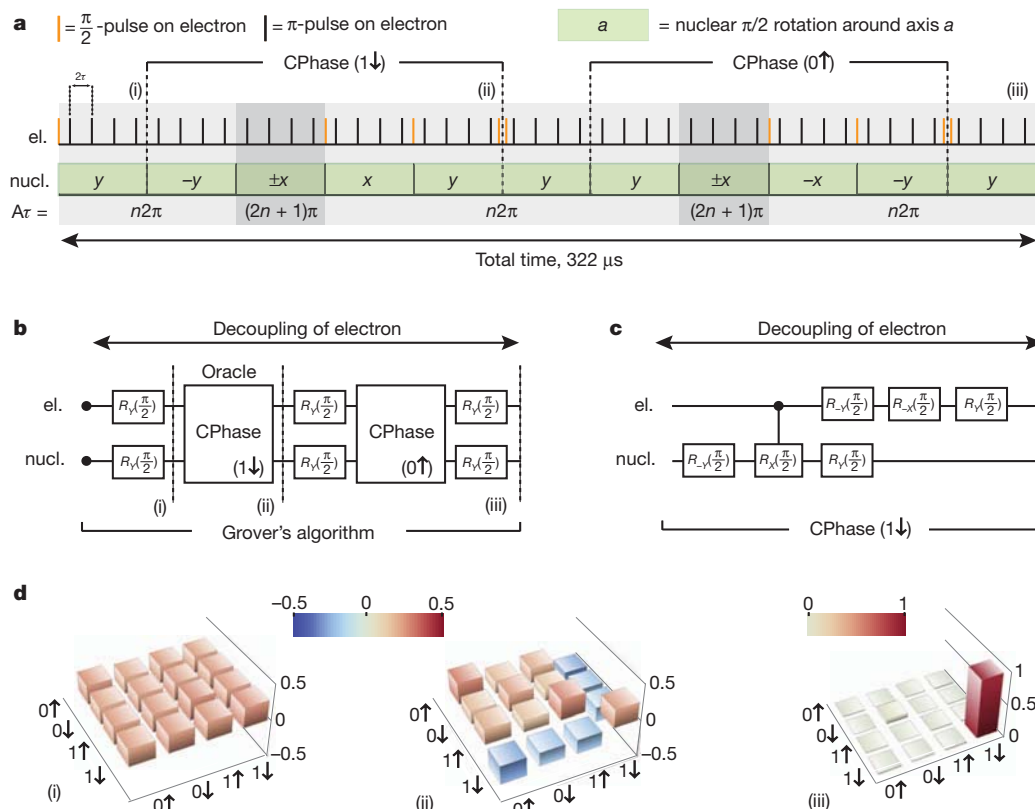
controllable manner, we inject low-frequency noise into the system<sup>23</sup>, thus shortening the electron spin-echo decay time,  $T_2$ , from 251(7) to 50(2)  $\mu\text{s}$  (Fig. 3a). Furthermore, we reduce the nuclear driving power such that the gate time becomes 120  $\mu\text{s}$ , more than twice  $T_2$ .

We first verify that the additional decoherence can be efficiently suppressed at the single-qubit level by dynamical decoupling. We observe that the electron spin coherence time increases from 50(2) to 234(8)  $\mu\text{s}$  as the number of the decoupling pulses is increased from one (spin echo) to sixteen (Fig. 3a), in agreement with previous studies<sup>4,5,7</sup>. Then we study the fidelity of the CNOT gate as a function of the number of decoupling pulses applied during its 120- $\mu\text{s}$  gate time. The resonance requirements for  $\tau$  are maintained by decreasing  $n$  as the number of pulses is increased (Fig. 3b). We apply the gate to the state  $(|0\rangle + i|1\rangle) \otimes |\uparrow\rangle$ ; this ideally yields the entangled state  $|\Psi^+\rangle = |0\rangle\uparrow + |1\rangle\downarrow$  (here  $|\uparrow\rangle$  and  $|\downarrow\rangle$  correspond to the nuclear spin states with  $m_I = +1$  and  $m_I = 0$ , respectively). Quantum state tomography reveals that the coherence of the output state, corresponding to the off-diagonal elements in the density matrix,  $\rho$ , grows rapidly with the number of pulses (Fig. 3c). The state fidelity<sup>1</sup>,  $F = \sqrt{\langle \Psi^+ | \rho | \Psi^+ \rangle}$ , reaches 96(1)% when there are 16 decoupling pulses (Fig. 3d), which corresponds to the overlap  $O = \langle \Psi^+ | \rho | \Psi^+ \rangle = 0.92(2)$ . Here the gate performs similarly to how it does without the introduced decoherence (Fig. 3d, dashed lines), showing that the gate efficiency remains high even in the regime where the gate time exceeds  $T_2$ . Moreover, comparison with single-qubit state fidelities under decoupling (taken from Fig. 3a) demonstrates that the electron spin coherence in the course of the gate operation is preserved as efficiently as it is for an idle electron

qubit (Fig. 3d). Numerical simulations provide evidence that the fidelity of the CNOT gate remains high even in the presence of much stronger decoherence caused by nitrogen impurities<sup>4,26</sup> (Supplementary Information), which is highly relevant for scalable quantum architectures in which chains of nitrogen atoms serve as data buses between nitrogen–vacancy centres<sup>21,22</sup>.

Finally we illustrate the power of the decoherence-protected gates by implementing a two-qubit algorithm in a hybrid solid-state spin register. We perform Grover's quantum algorithm<sup>1</sup> for searching for a given entry in an unstructured list of  $L$  elements. A classical search presents each entry in turn to an 'oracle', which outputs 1 if this is the target entry and outputs 0 otherwise. This requires  $\sim L/2$  oracle calls on average. The quantum algorithm encodes each entry as a state of an  $m$ -qubit system ( $2^m \geq L$ ) and presents a superposition of all entries to an oracle. The result is processed to increase the weight of the target state in the superposition. After of order  $\sqrt{L}$  iterations, this weight becomes close to 1. For two qubits ( $L = 4$ ), a single oracle call is enough to provide an exact answer.

Figure 4a shows all the pulses that implement the algorithm. Because the total execution time, 322  $\mu\text{s}$ , is almost 100 times longer than the electron spin dephasing time, high-fidelity execution of the algorithm is impossible without decoherence protection. The circuit diagram of the full computation is given in Fig. 4b for the target state  $|1\downarrow\rangle$ , with the conditional phase (CPhase) gate presented separately in Fig. 4c. Figure 4d shows quantum tomographic snapshots of the corresponding two-qubit states at different stages of the algorithm<sup>29</sup>. The fidelity of the resulting state is 95(1)%; for the other target states, the



**Figure 4 | Grover's search algorithm executed with decoherence-protected gates.** **a**, Pulse sequence implementing Grover's algorithm for the target state  $|1\downarrow\rangle$ . All nuclear spin rotations are implemented with decoherence-protected gates. Each nuclear  $\pi/2$  rotation is performed using two gate units. The electron and nuclear spin rotation axes are adjusted by changing the phase of the driving field. The total computation time is 322  $\mu\text{s}$ , which exceeds the electron spin-echo decay time,  $T_2 = 251 \mu\text{s}$ , by about 30%. **b**, Quantum circuit diagram of Grover's algorithm for the target state  $|1\downarrow\rangle$ , corresponding to the pulse

sequence shown in **a**. **c**, Quantum circuit diagram of the CPhase gate, implemented by embedding a CNOT gate between two nuclear  $\pi/2$ -rotation gates that rotate the nuclear spin basis. The pulses on the electron correct the single-qubit phase shift. **d**, Density matrices of the two-qubit system at different stages of Grover's algorithm. Stage (i) corresponds to the initial superposition of all two-qubit states, stage (ii) is the oracle's output and stage (iii) is the final state of the register; see also **a** and **b**. The fidelity of the final state is 95(1)% (overlap, 0.91(3)), mainly limited by longitudinal relaxation of the electron spin.

fidelities are 92(1)% ( $|0\uparrow\rangle$ ), 91(2)% ( $|0\downarrow\rangle$ ) and 91(1)% ( $|1\uparrow\rangle$ ) (Supplementary Information).

Our demonstration of decoherence-protected quantum gates based on resonantly applied decoupling pulses opens the way to high-fidelity transfer, processing and retrieval of quantum information in small quantum registers, which are critical tasks for future quantum repeaters and quantum computers<sup>1,2,13–16,30</sup>. Moreover, our gate design is compatible with quantum error correction and therefore marks an important step towards scalable, fault-tolerant quantum computation in a hybrid qubit architecture.

## METHODS SUMMARY

**System initialization and read-out.** All experiments are performed in a magnetic field of  $\sim 510$  G aligned parallel to the symmetry axis of the nitrogen–vacancy centre. The two-qubit system is initialized into the  $(m_s, m_I) = (0, +1)$  state by 4- $\mu$ s excitation using a green laser (wavelength, 532 nm; details in Supplementary Information). Electron spin polarization into the  $m_s = 0$  state is a result of a spin-dependent relaxation mechanism between the electronic excited state and electronic ground state. At 510 G, a level anticrossing in the electronic excited state allows electron–nuclear spin flip-flops, which, in combination with the mechanism responsible for electron spin polarization, leads to nuclear spin polarization into the  $m_I = +1$  state. The state of the system is read out by counting the spin-dependent number of photons emitted into the phonon sideband on green-laser excitation during a detection time window of 1–2  $\mu$ s (refs 27, 28; details in Supplementary Information).

**Influence of longitudinal relaxation of the electron spin.** All state fidelity numbers include errors due to spin relaxation within and out of the two-qubit subspace. From independently measured relaxation rates, we estimate that the probabilities to undergo relaxation out of and within the two-qubit subspace are respectively: 0.5% and 0.4% for Fig. 2d, 1% and 0.8% for Fig. 3, and 2.9% and 2.1% for Fig. 4 (Supplementary Information).

Received 6 October 2011; accepted 25 January 2012.

- Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, 2000).
- Ladd, T. D. *et al.* Quantum computers. *Nature* **464**, 45–53 (2010).
- Viola, L., Lloyd, S. & Knill, E. Universal control of decoupled quantum systems. *Phys. Rev. Lett.* **83**, 4888–4891 (1999).
- de Lange, G., Wang, Z. H., Ristè, D., Dobrovitski, V. V. & Hanson, R. Universal dynamical decoupling of a single solid-state spin from a spin bath. *Science* **330**, 60–63 (2010).
- Ryan, C. A., Hodges, J. S. & Cory, D. G. Robust decoupling techniques to extend quantum coherence in diamond. *Phys. Rev. Lett.* **105**, 200402 (2010).
- Barthel, C., Medford, J., Marcus, C. M., Hanson, M. P. & Gossard, A. C. Interlaced dynamical decoupling and coherent operation of a singlet-triplet qubit. *Phys. Rev. Lett.* **105**, 266808 (2010).
- Naydenov, B. *et al.* Dynamical decoupling of a single-electron spin at room temperature. *Phys. Rev. B* **83**, 081201 (2011).
- Bluhm, H. *et al.* Dephasing time of GaAs electron-spin qubits coupled to a nuclear bath exceeding 200  $\mu$ s. *Nature Phys.* **7**, 109–113 (2011).
- Bylander, J. *et al.* Noise spectroscopy through dynamical decoupling with a superconducting flux qubit. *Nature Phys.* **7**, 565–570 (2011).
- West, J. R., Lidar, D. A., Fong, B. H. & Gyure, M. F. High fidelity quantum gates via dynamical decoupling. *Phys. Rev. Lett.* **105**, 230503 (2010).
- Ng, K.-H., Lidar, D. A. & Preskill, J. Combining dynamical decoupling with fault-tolerant quantum computation. *Phys. Rev. A* **84**, 012305 (2011).
- Childress, L. *et al.* Coherent dynamics of coupled electron and nuclear spin qubits in diamond. *Science* **314**, 281–285 (2006).
- Childress, L., Taylor, J. M., Sorensen, A. S. & Lukin, M. D. Fault-tolerant quantum communication based on solid-state photon emitters. *Phys. Rev. Lett.* **96**, 070504 (2006).
- Morton, J. J. L. *et al.* Solid-state quantum memory using the  $^{31}\text{P}$  nuclear spin. *Nature* **455**, 1085–1088 (2008).
- Neumann, P. *et al.* Multipartite entanglement among single spins in diamond. *Science* **320**, 1326–1329 (2008).
- Jiang, L. *et al.* Repetitive readout of a single electronic spin via quantum logic with nuclear spin ancillae. *Science* **326**, 267–272 (2009).
- Kubo, Y. *et al.* Strong coupling of a spin ensemble to a superconducting resonator. *Phys. Rev. Lett.* **105**, 140502 (2010).
- Rabl, P. *et al.* A quantum spin transducer based on nanoelectromechanical resonator arrays. *Nature Phys.* **6**, 602–608 (2010).
- Fuchs, G. D., Burkard, G., Klimov, P. & Awschalom, D. D. A quantum memory intrinsic to single nitrogen–vacancy centres in diamond. *Nature Phys.* **7**, 789–793 (2011).
- Arcizet, O. *et al.* A single nitrogen-vacancy defect coupled to a nanomechanical oscillator. *Nature Phys.* **7**, 879–883 (2011).
- Cappellaro, P., Viola, L. & Ramanathan, C. Coherent-state transfer via highly mixed quantum spin chain. *Phys. Rev. A* **83**, 032304 (2011).
- Yao, N. Y. *et al.* Scalable architecture for a room temperature solid-state quantum processor. *Nature Commun.* (in the press); preprint at (<http://arxiv.org/abs/1012.2864>) (2010).
- Biercuk, M. J. *et al.* Optimized dynamical decoupling in a model quantum memory. *Nature* **458**, 996–1000 (2009).
- Sagi, Y., Almog, I. & Davidson, N. Process tomography of dynamical decoupling in a dense cold atomic ensemble. *Phys. Rev. Lett.* **105**, 053201 (2010).
- Du, J. *et al.* Preserving electron spin coherence in solids by optimal dynamical decoupling. *Nature* **461**, 1265–1268 (2009).
- Hanson, R., Dobrovitski, V. V., Feiguin, A. E., Gywat, O. & Awschalom, D. D. Coherent dynamics of a single spin interacting with an adjustable spin bath. *Science* **320**, 352–355 (2008).
- Smeltzer, B., McIntyre, J. & Childress, L. Robust control of individual nuclear spins in diamond. *Phys. Rev. A* **80**, 050302 (2009).
- Steiner, M., Neumann, P., Beck, J., Jelezko, F. & Wrachtrup, J. Universal enhancement of the optical readout fidelity of single electron spins at nitrogen-vacancy centers in diamond. *Phys. Rev. B* **81**, 035205 (2010).
- DiCarlo, L. *et al.* Demonstration of two-qubit algorithms with a superconducting quantum processor. *Nature* **460**, 240–244 (2009).
- Bermudez, A., Jelezko, F., Plenio, M. B. & Retzker, A. Electron-mediated nuclear-spin interactions between distant NV centers. *Phys. Rev. Lett.* **107**, 150503 (2011).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank L. DiCarlo, F. Jelezko, M. D. Lukin and L. M. K. Vandersypen for discussions and comments. T.v.d.S., H.B. and R.H. acknowledge support from the Dutch Organization for Fundamental Research on Matter and the Netherlands Organization for Scientific Research. D.D.A. acknowledges support from DARPA QuEST, AFOSR and ARO MURI, and R.H. acknowledges support from DARPA QuEST. D.A.L. was sponsored by the National Science Foundation under grant numbers CHM-924318, CHM-1037992 and PHY-0969969, ARO MURI grant W911NF-11-1-0268, and by the US Department of Defense. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressly or implied, of the US Government. Work at Ames Laboratory was supported by the Department of Energy, Basic Energy Sciences under contract number DE-AC02-07CH11358.

**Author Contributions** Z.H.W., D.A.L. and V.V.D. designed the gate and did the theoretical analysis. H.B. and D.M.T. made the device. T.v.d.S., M.S.B., T.H.T., D.D.A. and R.H. designed and performed the experiments. T.v.d.S., R.H. and V.V.D. wrote the manuscript. All authors discussed the results and commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to V.V.D. ([slava@ameslab.gov](mailto:slava@ameslab.gov)).



# Past extreme warming events linked to massive carbon release from thawing permafrost

Robert M. DeConto<sup>1</sup>, Simone Galeotti<sup>2</sup>, Mark Pagani<sup>3</sup>, David Tracy<sup>1</sup>, Kevin Schaefer<sup>4</sup>, Tingjun Zhang<sup>4,7</sup>, David Pollard<sup>5</sup> & David J. Beerling<sup>6</sup>

Between about 55.5 and 52 million years ago, Earth experienced a series of sudden and extreme global warming events (hyperthermals) superimposed on a long-term warming trend<sup>1</sup>. The first and largest of these events, the Palaeocene–Eocene Thermal Maximum (PETM), is characterized by a massive input of carbon, ocean acidification<sup>2</sup> and an increase in global temperature of about 5 °C within a few thousand years<sup>3</sup>. Although various explanations for the PETM have been proposed<sup>4–6</sup>, a satisfactory model that accounts for the source, magnitude and timing of carbon release at the PETM and successive hyperthermals remains elusive. Here we use a new astronomically calibrated cyclostratigraphic record from central Italy<sup>7</sup> to show that the Early Eocene hyperthermals occurred during orbits with a combination of high eccentricity and high obliquity. Corresponding climate–ecosystem–soil simulations accounting for rising concentrations of background greenhouse gases<sup>8</sup> and orbital forcing show that the magnitude and timing of the PETM and subsequent hyperthermals can be explained by the orbitally triggered decomposition of soil organic carbon in circum-Arctic and Antarctic terrestrial permafrost. This massive carbon reservoir had the potential to repeatedly release thousands of petagrams (10<sup>15</sup> grams) of carbon to the atmosphere–ocean system, once a long-term warming threshold had been reached just before the PETM. Replenishment of permafrost soil carbon stocks following peak warming probably contributed to the rapid recovery from each event<sup>9</sup>, while providing a sensitive carbon reservoir for the next hyperthermal<sup>10</sup>. As background temperatures continued to rise following the PETM, the areal extent of permafrost steadily declined, resulting in an incrementally smaller available carbon pool and smaller hyperthermals at each successive orbital forcing maximum. A mechanism linking Earth's orbital properties with release of soil carbon from permafrost provides a unifying model accounting for the salient features of the hyperthermals.

The PETM is characterized by a shift of >3‰ in the stable carbon isotopic composition ( $\delta^{13}\text{C}$ ) of the surficial carbon pool, implying a huge injection of  $^{13}\text{C}$ -depleted carbon into the ocean and atmosphere. The source of this carbon is commonly attributed to the release of  $^{13}\text{C}$ -depleted methane from the destabilization of marine methane clathrates<sup>6</sup>. The isotopic composition of this source ( $\delta^{13}\text{C} = -60\text{‰}$ ) provides a reasonable explanation for the isotope shift, but implies a carbon release (of ~2,000 PgC) that is both inadequate to explain the magnitude of observed warming<sup>11,12</sup> and difficult to justify in clathrate modelling studies<sup>13</sup>. Other proposed carbon sources include a comet impact<sup>14</sup>, a global conflagration of peat deposits<sup>5</sup>, thermogenic CO<sub>2</sub> and CH<sub>4</sub> production in the North Atlantic<sup>15</sup>, and the desiccation of organic matter in shallow continental seaways<sup>4</sup>. However, these suppositions fail to explain the sequence of multiple, progressively smaller hyperthermals that occurred over a three-million-year interval following the PETM<sup>1,16</sup>. The possibility that these events correspond to similar

orbital geometries<sup>17</sup> suggests they share a common trigger mechanism and are not driven by rare events, tectonics or stochastic processes.

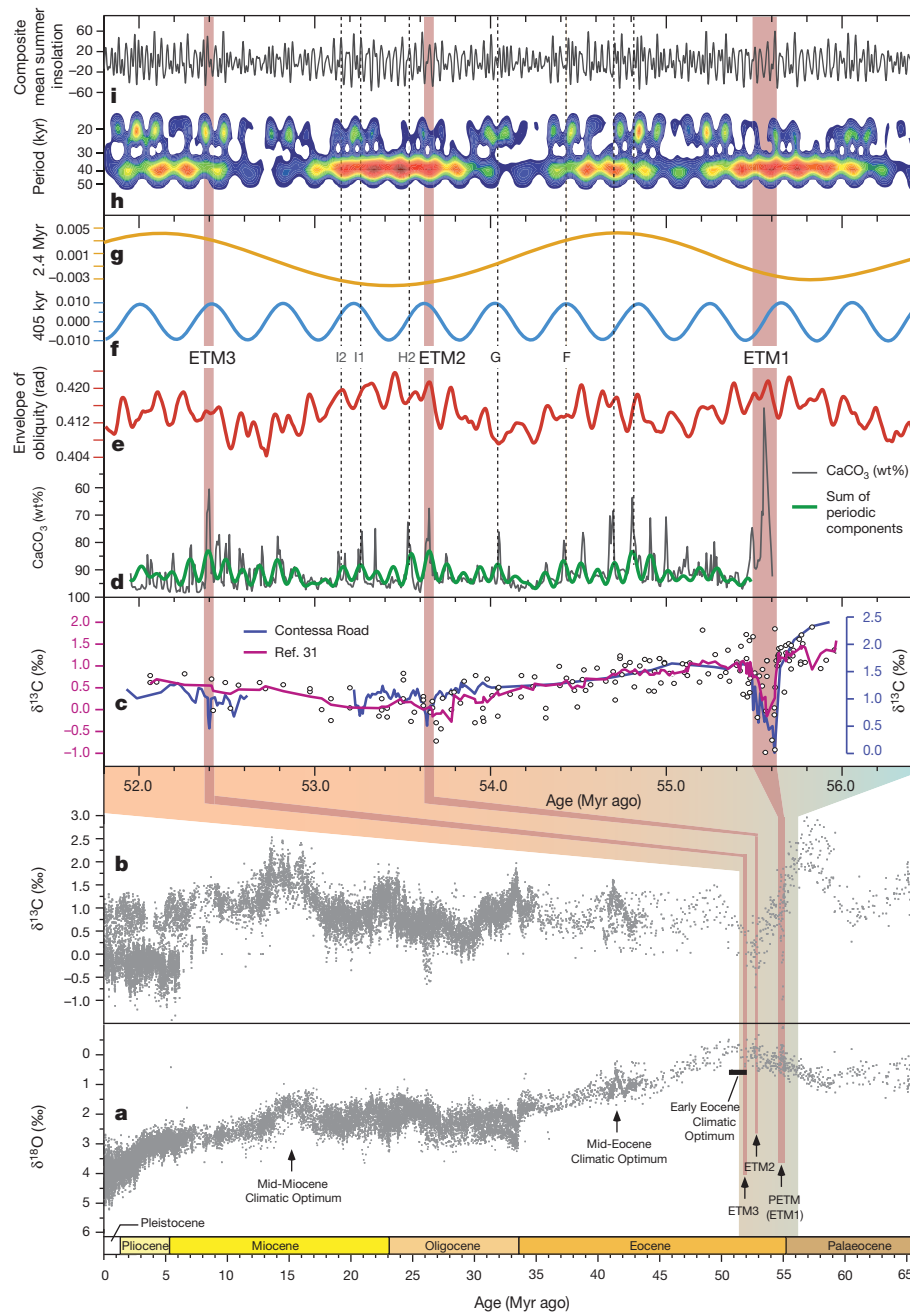
We investigated the relationship between orbital forcing and the timing of hyperthermals by determining their astrochronological age in the Contessa Road outcrop near Gubbio, Italy (Fig. 1 and Supplementary Information). The evolutive spectrum (obtained using wavelet analysis) of mean summer insolation at 75 °S<sup>18</sup> shows maxima at irregular intervals resulting from the combination of the long-term modulation of orbital eccentricity (2.4 Myr) and obliquity (1.2 Myr). Peaks in high-latitude insolation at concomitant maxima in the long (405-kyr) eccentricity and (1.2-Myr) obliquity bands in Cenozoic astronomical solutions<sup>18</sup> correspond with the onset of the major hyperthermal events at ~55.7 Myr ago (the PETM, also known as Eocene Thermal Maximum 1, ETM1), ~53.7 Myr ago (ETM2) ~52.4 Myr ago (ETM3), and other smaller events recorded in different settings<sup>16,19</sup>. Uncertainties in orbital solutions older than ~40 Myr (ref. 18) preclude the unequivocal orbital tuning of specific Eocene events. However, the floating cyclochronology at Contessa Road allows long-term modulation patterns to be determined, regardless of absolute ages. In our analysis, PETM (ETM1), ETM2 and ETM3 all align with obliquity and eccentricity maxima irrespective of which absolute age model is used<sup>7,19</sup>. This coincidence of hyperthermals with combined high-eccentricity and high-obliquity forcing points to the influence of polar latitudes, where seasonal insolation is strongly affected by obliquity and precession during high-eccentricity orbits.

We propose that the hyperthermals are linked to high-latitude orbital forcing through carbon-cycle feedbacks involving permafrost soil carbon (PFSC). The soil carbon reservoir has been dismissed as too small to cause an event like the PETM<sup>4</sup>, but this view overlooks the potential for large PFSC stocks on Antarctica, before the continent became glaciated around 34 Myr ago. Carbon is frozen into permafrost through aeolian deposition, alluvial sedimentation and vertical peat deposition, which slowly increase soil depth on timescales of decades to millennia<sup>20</sup>. As sedimentation increases soil depth, roots and organic material at the bottom of the active layer become frozen into the permafrost, sequestering this organic matter from the active carbon cycle. A new reconstruction of Antarctic palaeogeography<sup>21</sup> shows that West Antarctica was mostly subaerial during the Palaeogene, increasing the continent's surface area by ~25% relative to today. Assuming that Antarctica remained mostly ice-free through the late Palaeocene and early Eocene, most of its  $12.1 \times 10^6 \text{ km}^2$  land surface would have been covered by forest and tundra (Fig. 2a), with the potential for storing massive amounts of permafrost carbon. Moreover, the continent's polar position would have made these frozen carbon stocks sensitive to the high-obliquity and high-eccentricity orbital forcing corresponding to the onset of the hyperthermals.

We tested the sensitivity of polar ecosystems and permafrost to increasing greenhouse-gas (GHG) concentrations<sup>8</sup>, high-obliquity

<sup>1</sup>Department of Geosciences, University of Massachusetts, Amherst, Massachusetts 01002, USA. <sup>2</sup>Earth, Life, and Environmental Sciences Department, University of Urbino, 61029 Urbino, Italy.

<sup>3</sup>Department of Geology and Geophysics, Yale University, New Haven, Connecticut 06520, USA. <sup>4</sup>National Snow and Ice Data Center, Cooperative Institute for Research in Environmental Sciences, University of Colorado at Boulder, Boulder, Colorado 80309, USA. <sup>5</sup>Earth and Environmental Systems Institute, Pennsylvania State University, University Park, Pennsylvania 16802, USA. <sup>6</sup>Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK. <sup>7</sup>Ministry of Education Key Laboratory of West China's Environmental System, Lanzhou University, 222 Tianshuinanlu, Lanzhou, Gansu 730000, China.

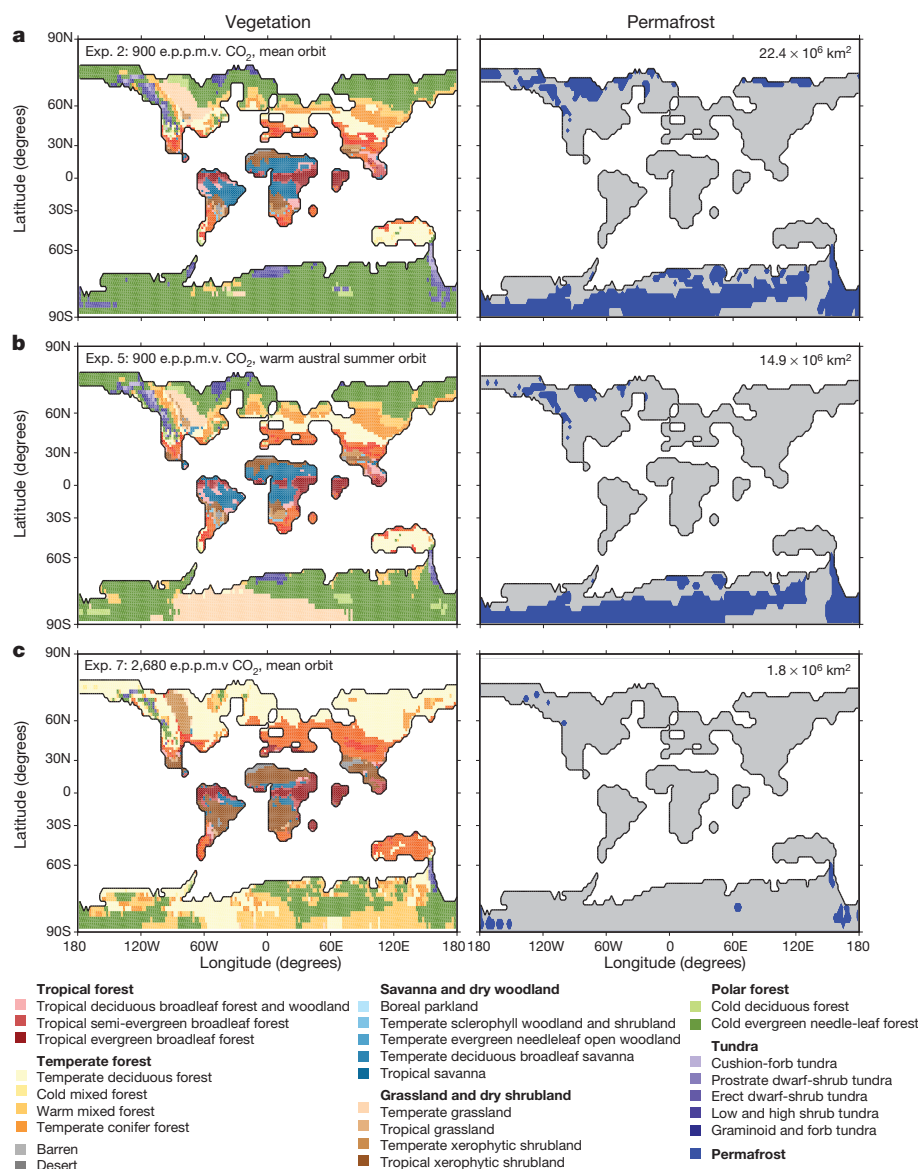


**Figure 1 | Orbital phasing of Early Eocene hyperthermals.** **a, b,** The Cenozoic record of benthic oxygen (**a**) and carbon (**b**) isotope values<sup>31</sup> shows the late Palaeocene–early Eocene gradual warming trend (shaded blue to red) that culminates at the Early Eocene Climatic Optimum. Positions of hyperthermals PETM (ETM1), ETM2 and ETM3 are shown. Panels above show the shaded area in more detail. **c,** Major early Eocene hyperthermals recorded in the bulk carbon isotope record from Contessa Road (blue line) relative to the benthic carbon isotope stack in ref. 31 (dots and pink line). **d,**  $\text{CaCO}_3$  record from the Contessa Road section (grey line) and MTM reconstruction of individual components exceeding the 99% confidence level (green line). **e,** Envelope (Hilbert transform) of the obliquity parameter from the La04 astronomical solution<sup>18</sup>. **f,** The long-term (405 kyr) modulation component of eccentricity from La04<sup>18</sup>. **g,** The very long-term (~2.4 Myr)

modulation component of eccentricity from La04<sup>18</sup>. The astrochronological age of PETM (ETM1), ETM2 and ETM3 (grey bands) is based on Contessa Road. The inferred positions of other minor dissolution events and carbon isotope excursions (CIE; dashed lines), including F, G, H2, I1, I2 and the two unnamed events between 54.6 and 54.8 Myr ago<sup>16,17,19</sup>, are also shown. All major events (ETMs) occur at maximum values of obliquity and minimal  $\text{CaCO}_3$  values, and mimic the long-term modulation of obliquity. **h,** Evulsive spectrum of combined mean summer insolation (CMSI) at high latitudes shows intervals of maximum power in the higher frequency components (precession and obliquity) across hyperthermals. **i,** The CMSI series obtained by summing mean summer insolation at 65° N (21 June to 21 September) and 75° S (21 December to 21 March). The 65° N insolation values are doubled to account for the larger land area of the Northern Hemisphere (see Supplementary Information).

and high-eccentricity orbital forcing, and underlying long-term warming between ~62 and 52 Myr ago (Fig. 1) using a Global Climate Model (GCM) with coupled terrestrial biosphere and soil components (Methods). The simulations (Table 1) account for Early Eocene palaeogeography (Supplementary Information) with a larger, ice-free Antarctica<sup>21</sup> and progressively increasing  $\text{CO}_2$ ,  $\text{CH}_4$  and  $\text{N}_2\text{O}$

atmospheric mixing ratios<sup>8</sup>. Four different prescribed orbits were applied at each GHG level (Table 1). The orbits include a mean orbital configuration with moderate obliquity and zero eccentricity (no precession forcing), a simulation with high obliquity and zero eccentricity, and two simulations with combined high obliquity and high eccentricity, corresponding to the onset of the hyperthermals (Fig. 1). In simulations



**Figure 2 | Climate-biome and permafrost simulations in response to increasing background GHG levels and orbital forcing.** Left, simulated biomes (corresponding vegetation types are shown below); right, permafrost (blue). CO<sub>2</sub> concentrations in equivalent parts per million by volume (e.p.p.m.v.) are shown at the top left of each panel and total permafrost area is shown at top right. Panels a–c represent a scenario of gradually increasing

background GHG levels, leading to experiment 2 (Table 1) at 900 e.p.p.m.v. CO<sub>2</sub> (a), initial permafrost thaw and carbon release triggered by high-obliquity and high-eccentricity orbital forcing (b; corresponds to experiment 5 in Table 1), and enough carbon mobilized to increase GHG concentrations above 2,680 e.p.p.m.v. CO<sub>2</sub>, causing 6 °C warming (c; corresponds to experiment 7 in Table 1).

**Table 1 | Climate-biome simulations, relevant inputs, and permafrost areas**

Exp.	*CO <sub>2</sub> (e.p.p.m.v.)	CO <sub>2</sub> (p.p.m.v.)	CH <sub>4</sub> (p.p.b.v.)	N <sub>2</sub> O (p.p.b.v.)	Ecc.	Prec. (°)	Obl. (°)	MAT (°C)	NH PF (10 <sup>6</sup> km <sup>2</sup> )	SH PF (10 <sup>6</sup> km <sup>2</sup> )	Tot. PF (10 <sup>6</sup> km <sup>2</sup> )
Cont.	400	355	1714	311	0.017	77.06	23.765	14.65	23.72	NA	23.72
1	550	500	1200	350	0.00	0.00	23.5	18.31	19.47	11.77	31.24
<b>2</b>	<b>900</b>	<b>750</b>	<b>2100</b>	<b>375</b>	<b>0.00</b>	<b>0.00</b>	<b>23.5</b>	<b>20.30</b>	<b>12.04</b>	<b>10.31</b>	<b>22.35</b>
3	900	750	2100	375	0.00	0.00	24.5	20.58	7.24	9.70	16.94
4	900	750	2100	375	0.05	270.00	24.5	20.63	5.82	9.70	15.52
<b>5</b>	<b>900</b>	<b>750</b>	<b>2100</b>	<b>375</b>	<b>0.05</b>	<b>90.00</b>	<b>24.5</b>	<b>20.92</b>	<b>5.77</b>	<b>9.21</b>	<b>14.98</b>
6	1,275	1,000	3000	400	0.00	0.00	23.5	22.39	1.59	8.27	9.86
<b>7</b>	<b>2,680</b>	<b>2,000</b>	<b>3500</b>	<b>450</b>	<b>0.00</b>	<b>0.00</b>	<b>23.5</b>	<b>26.30</b>	<b>0.45</b>	<b>1.33</b>	<b>1.78</b>
8	5,360	4,000	3500	450	0.00	0.00	23.5	30.45	NA	NA	NA

Experiments (Exp.) 2, 5 and 7 (bold font) correspond to the warming/orbital sequence in Fig. 2. \*CO<sub>2</sub> is the equivalent volume mixing ratio (e.p.p.m.v.) of each simulation, if only CO<sub>2</sub> were increased and CH<sub>4</sub> and N<sub>2</sub>O were held at their pre-industrial levels of 700 and 270 p.p.b.v., respectively. Ecc., eccentricity. Prec., precession (the prograde angle from perihelion to the vernal equinox). Obl., obliquity (that is, tilt). Simulations with zero eccentricity and obliquity at 23.5° (exp. 1, 2, 6, 7, 8) correspond to the mean orbit simulations discussed in the text. MAT, simulated global-mean annual surface (2 m) air temperature. Permafrost (PF) areas in the Northern Hemisphere (NH) and the Southern Hemisphere (Antarctica; SH), and the total permafrost area (Tot.) are given. An equilibrated modern control simulation (Cont.) is shown for comparison with observations. Experiment 2 provides the pre-PETM baseline climate for our carbon calculations, as supported by terrestrial data-model comparisons (Supplementary Information). NA, not applicable.



with combined high eccentricity and high obliquity, precession is varied to place perihelion in either January or July, producing summers with high-intensity insolation over one pole and summers with lower insolation intensity but longer duration over the other.

Simulations with 550–900 effective parts per million by volume (e.p.p.m.v.) CO<sub>2</sub> (Table 1) representing pre-PETM (late Palaeocene) conditions show polar latitudes dominated by tundra, evergreen forest, and permafrost—environments well suited to sequestering PFSC<sup>22</sup> (Fig. 2a). At 900 e.p.p.m.v. CO<sub>2</sub>, global-mean surface temperature is 6 °C warmer than today and Antarctic summers are too warm to allow glaciation; but  $22.4 \times 10^6$  km<sup>2</sup> of permafrost (roughly equivalent to the total modern inventory) remains in the high latitudes of both hemispheres. The 900 e.p.p.m.v. CO<sub>2</sub> simulation (experiment 2) compares favourably with terrestrial estimates of late Palaeocene mean annual temperature (MAT) over a range of latitudes (Supplementary Information), indicating that it broadly represents late Palaeocene climate with substantial permafrost in the higher terrain of both poles.

In a framework of gradual warming, annual surface heat budgets in the permafrost zone approach net-zero as GHG levels reach 900 e.p.p.m.v. CO<sub>2</sub>, making permafrost stability highly sensitive to orbital forcing once this threshold is reached. Relative to the simulation with a mean orbit (experiment 2), the combination of high obliquity and high eccentricity increases summer insolation by more than 85 W m<sup>-2</sup> over one pole, while increasing summer duration over the other (Supplementary Information). With high eccentricity, total permafrost area is most sensitive to a warm austral summer orbit (experiment 5), but both hemispheres lose permafrost simultaneously regardless of the timing of perihelion (Table 1), because permafrost is sensitive to both high-intensity and long-duration summers. At 900 e.p.p.m.v. CO<sub>2</sub>, the combination of high-eccentricity and high-obliquity orbital forcing reduces the global permafrost region by  $7.4 \times 10^6$  km<sup>2</sup> (33%). Simulations with high obliquity and zero eccentricity also show reductions in permafrost due to increased seasonality, but the combined effect of high obliquity and high eccentricity is greater (Table 1), with the potential to thaw vast areas of permafrost once a warming threshold is reached.

We estimate the response of the soil carbon reservoir to these orbitally driven changes in permafrost area by calculating the available PFSC reservoir in the areas that undergo thaw. Lacking details of Palaeocene–Eocene topography and observations of long-term permafrost depositional processes (especially in the Antarctic interior), we calculate the available PFSC in each simulation by assuming that past permafrost conditions were similar to today (Methods). A range of observed values of fractional wetland area, carbon density, and carbon deposit thickness in the permafrost zone are used to produce an ensemble of plausible PFSC inventories, providing a best estimate of the reservoir in each simulation and an estimate of uncertainty (Table 2). The PFSC reservoir in our modern control simulation (1,526 Pg C) is close to observations (1,672 Pg C)<sup>22</sup>, but our Palaeogene values are probably underestimated, because they assume that the fraction of wetland, ancient carbon densities, and vertical accumulations in the permafrost zone are comparable to those in analogous modern settings. In contrast, Palaeocene–Eocene wetlands

are estimated to have been ~3 times more extensive than today<sup>8,23,24</sup>, as supported by an analysis of the ancient topography (Supplementary Information), indicating a greater fractional land area capable of supporting peat formation. Furthermore, unlike the Quaternary, the polar regions of the early Cenozoic remained undisturbed by ice sheets for millions of years, with the potential for >10<sup>6</sup> years of continuous carbon deposition before the PETM and between major hyperthermals (Fig. 1). Hence, more extensive and thicker PFSC deposits than those considered here are probable, as evidenced by >10-m-thick Palaeogene coal deposits in some polar locations<sup>25</sup>. Such deposits could also account for trends in marine sulphur and carbon isotopes, indicating massive terrestrial carbon burial before the PETM<sup>5</sup>. Regardless, the conservative approach used here still points to a massive and vulnerable terrestrial carbon pool on a world free of major ice sheets.

Assuming that experiment 2 (Fig. 2a) broadly represents late Palaeocene climate, our best estimate of the pre-PETM PFSC carbon pool is  $3,728 \pm 1,033$  Pg C (Table 2), nearly half of which is in Antarctica. This estimate more than doubles if we assume that the permafrost region in the wetter Palaeogene was dominated by peatland. After reaching a warming threshold at ~900 e.p.p.m.v. CO<sub>2</sub>, the combination of high-eccentricity and high-obliquity orbital forcing releases more than 1,200 Pg C to the atmosphere. Perihelion during austral summer (experiment 5) causes the greatest carbon release ( $1,237 \pm 331$  Pg C), although perihelion during boreal summer (experiment 4) has a nearly equivalent effect. This initial carbon input at rates up to ~1.5 Pg C yr<sup>-1</sup> has the potential to raise atmospheric CO<sub>2</sub> by more than 550 p.p.m.v. (see Methods), providing additional radiative forcing, more warming and further permafrost loss. As illustrated by the sequence of simulations in Fig. 2, this orbitally triggered permafrost–carbon–warming feedback liberates almost the entire global PFSC reservoir, releasing  $3,434 \pm 951$  Pg C within the observed 10<sup>4</sup>-year timescale of carbon release. This raises global-mean temperature in our model by 6 °C (Table 1), accounting for the warming at the PETM (ETM1) without invoking high climate sensitivity to CO<sub>2</sub> or additional feedbacks involving other carbon reservoirs<sup>6</sup>. Furthermore, our estimates of carbon release from the PFSC reservoir are broadly consistent with geochemical model constraints<sup>11,12</sup> and provide an explanation for a previously unexplained pulse of organic carbon into the Southern Ocean at the PETM (ETM1)<sup>26</sup>.

These results show the potential for high-latitude climate forcing to trigger massive terrestrial carbon release, initiating positive warming feedbacks that can account for the sudden and extreme nature of past hyperthermals. We find that Antarctica and high elevations of the circum-Arctic were suitable locations for massive carbon storage during the late Palaeocene and Early Eocene. During long-term warming, these environments eventually reached a climatic threshold, at which permafrost thaw and sudden release of stored soil carbon<sup>27</sup> were triggered during maxima in the long eccentricity and obliquity orbital cycles (Supplementary Fig. 1). This model calls on the direct release of CO<sub>2</sub> from the decomposition of terrestrial PFSC reserves, although background fluxes of CH<sub>4</sub> and N<sub>2</sub>O were also probably enhanced<sup>8,28</sup>, possibly contributing additional warming through atmospheric water vapour and cloud feedbacks<sup>24</sup>. Following the exhaustion of the PFSC

**Table 2 | Simulated permafrost soil carbon**

Exp.	*CO <sub>2</sub> (e.p.p.m.v.)	NH PFSC (Pg C)	Antarctic PFSC (Pg C)	Global PFSC (Pg C)	Release (Pg C)
Cont.	400	1526	0	1,526	NA
1	550	3,268 ± 901	1,976 ± 545	5,245 ± 1,446	NA
<b>2</b>	<b>900</b>	<b>2,008 ± 557</b>	<b>1,720 ± 477</b>	<b>3,728 ± 1,033</b>	<b>NA</b>
3	900	962 ± 269	1,604 ± 448	2,800 ± 782	929 ± 253
4	900	964 ± 267	1,538 ± 426	2,591 ± 717	1,138 ± 316
<b>5</b>	<b>900</b>	<b>1,203 ± 334</b>	<b>1,612 ± 448</b>	<b>2,491 ± 692</b>	<b>1,237 ± 341</b>
6	1,275	264 ± 73	1,372 ± 382	1,636 ± 455	2,093 ± 578
<b>7</b>	<b>2,680</b>	<b>74 ± 21</b>	<b>221 ± 62</b>	<b>295 ± 82</b>	<b>3,434 ± 951</b>
8	5,360	0	0	0	3,728 ± 1,033

Northern Hemisphere (NH), Antarctic, and global total permafrost soil carbon (PFSC) in each simulation are ensemble mean values with the standard deviation representing uncertainty (Methods). Subtracting the total PFSC in each simulation from the pre-PETM simulation (exp. 2) provides an estimate of carbon release (right hand column) through the warming sequence in Fig. 2. Bold font has same meaning as in Table 1.

reservoir at the onset of each hyperthermal, extreme warmth and high precipitation rates would have increased silicate weathering, contributing to some CO<sub>2</sub> drawdown within the first few 10<sup>3</sup>–10<sup>4</sup> years of the event. Once initial cooling and attenuation of orbital variability allowed permafrost to become re-established at high elevations, increasing PFSC sequestration would have accelerated CO<sub>2</sub> drawdown despite the diminishing effectiveness of silicate weathering in a cooling climate<sup>29</sup>. On the basis of Holocene rates of PFSC sequestration<sup>22</sup>, this mechanism can account for the rapid (10<sup>4</sup>-year) recovery from the hyperthermals evidenced in carbon isotope records, which has been difficult to explain through silicate weathering alone<sup>9</sup>. The diminishing magnitude of successive Eocene hyperthermals superposed on a backdrop of long-term warming is also explained, given progressively smaller available carbon inventories (Table 1) as permafrost became increasingly restricted to higher elevations.

## METHODS SUMMARY

Model simulations use an updated version of the GENESIS v. 3.0 GCM inter-actively coupled to the BIOME4 equilibrium vegetation model. The atmospheric component is coupled to surface models including a 50-m slab ocean model with predicted sea surface temperatures, diffusive heat transport, and dynamic-thermodynamic sea ice. Terrestrial land surface components include multi-layer snow and soil models and potential Eocene vegetation distributions are predicted. Model details, early Eocene boundary conditions, validation, and aspects of the soil component relevant to permafrost are discussed in Methods and Supplementary Information. A suite of Palaeocene–Eocene simulations (Table 1) was run with progressively increasing levels of CO<sub>2</sub>, CH<sub>4</sub> and N<sub>2</sub>O volume mixing ratios<sup>8</sup>, and a range of orbital configurations, including high-eccentricity and high-obliquity orbits corresponding to the onset of hyperthermals (Fig. 1).

PFSC calculations (Table 2) divide permafrost carbon into peatland, near surface, and deep pools such that PFSC =  $A_{\text{tot}} R \rho D$ , where  $A_{\text{tot}}$  is the permafrost area from each simulation,  $R$  is the area ratio,  $\rho$  is permafrost carbon density, and  $D$  is the thickness of the PFSC deposit. Values of  $R$ ,  $\rho$  and  $D$  for peatland, near surface, and deep pools are based on theoretical constraints<sup>30</sup> and the modern PFSC inventory<sup>22</sup>, resulting in a 12-member ensemble. The ensemble mean is our best estimate of frozen PFSC and the standard deviation represents uncertainty. Values of  $R$ ,  $\rho$  and  $D$  are described in Methods and Supplementary Information. Carbon releases (Table 2) assume that experiment 2 represents pre-PETM conditions.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 24 April 2011; accepted 2 February 2012.**

- Nicolo, M. J., Dickens, G. R., Hollis, C. J. & Zachos, J. Multiple early Eocene hyperthermals: their sedimentary expression on the New Zealand continental margin and in the deep sea. *Geology* **35**, 699–702 (2007).
- Zachos, J. C. *et al.* Rapid acidification of the ocean during the Paleocene–Eocene Thermal Maximum. *Science* **308**, 1611–1615 (2005).
- Sluijs, A. *et al.* Subtropical Arctic Ocean temperatures during the Palaeocene/Eocene thermal maximum. *Nature* **441**, 610–613 (2006).
- Higgins, J. A. & Schrag, D. Beyond methane: towards a theory for the Paleocene–Eocene Thermal Maximum. *Earth Planet. Sci. Lett.* **245**, 523–537 (2006).
- Kurtz, A. C., Kump, L. R., Arthur, M. A., Zachos, J. C. & Paytan, A. Early Cenozoic decoupling of the global carbon and sulfur cycles. *Paleoceanography* **18**, 1090, <http://dx.doi.org/10.1029/2003PA000908> (2003).
- Dickens, G. R., O’Neil, J. R., Rea, D. K. & Owen, R. M. Dissociation of oceanic methane hydrate as a cause of the carbon isotope excursion at the end of the Palaeocene. *Paleoceanography* **10**, 965–971 (1995).
- Galeotti, S. *et al.* Orbital chronology of Early Eocene hyperthermals from the Contessa Road section, central Italy. *Earth Planet. Sci. Lett.* **290**, 192–200 (2010).
- Beerling, D. J., Fox, A., Stevenson, D. S. & Valdes, P. J. Enhanced chemistry-climate feedbacks in past greenhouse worlds. *Proc. Natl Acad. Sci. USA* **108**, 9770–9775 (2011).
- Bowen, G. J. & Zachos, J. Rapid carbon sequestration at the termination of the Palaeocene–Eocene Thermal Maximum. *Nature Geosci.* **3**, 866–869 (2010).
- Dickens, J. Rethinking the global carbon cycle with a large, dynamic and microbially mediated gas hydrate capacitor. *Earth Planet. Sci. Lett.* **213**, 169–183 (2003).

- Zeebe, R. E., Zachos, J. & Dickens, G. R. Carbon dioxide forcing alone insufficient to explain Palaeocene–Eocene Thermal Maximum warming. *Nature Geosci.* **2**, 576–580 (2009).
- Panchuk, K., Ridgwell, A. & Kump, L. R. Sedimentary response to Paleocene–Eocene Thermal Maximum carbon release: a model-data comparison. *Geology* **36**, 315–318 (2008).
- Buffett, B. & Archer, D. Global inventory of methane clathrate: sensitivity to changes in the deep ocean. *Earth Planet. Sci. Lett.* **227**, 185–199 (2004).
- Kent, D. V. *et al.* A case for a comet impact trigger for the Paleocene/Eocene thermal maximum and carbon isotope excursion. *Earth Planet. Sci. Lett.* **211**, 13–26 (2003).
- Storey, M., Duncan, R. A. & Swisher, C. C. III. Paleocene–Eocene Thermal Maximum and the opening of the northeast Atlantic. *Science* **316**, 587–589 (2007).
- Cramer, B. S., Wright, J. D., Kent, D. V. & Aubry, M.-P. Orbital climate forcing of  $\delta^{13}\text{C}$  excursions in the late Paleocene–early Eocene (chrons C24n–C25n). *Paleoceanography* **18**, 1097, <http://dx.doi.org/10.1029/2003PA000909> (2003).
- Lourens, L. *et al.* Astronomical pacing of late Palaeocene to early Eocene hyperthermal events. *Nature* **435**, 1083–1087 (2005).
- Laskar, J. *et al.* A long-term numerical solution for the insolation quantities of the Earth. *Astron. Astrophys.* **428**, 261–285 (2004).
- Westerhold, T., Röhl, U., McCarren, H. K. & Zachos, J. C. Latest on the absolute age of the Paleocene–Eocene Thermal Maximum (PETM): new insights from exact stratigraphic position of key ash layers +19 and –17. *Earth Planet. Sci. Lett.* **287**, 412–419 (2009).
- Schuur, E. A. G. *et al.* Vulnerability of permafrost carbon to climate change: implications for the global carbon cycle. *Bioscience* **58**, 701–714 (2008).
- Wilson, D. S. & Luyendyk, B. P. West Antarctic paleotopography estimated at the Eocene–Oligocene climate transition. *Geophys. Res. Lett.* **36**, L16302 (2009).
- Tarnocai, C. *et al.* Soil organic carbon pools in the northern circumpolar permafrost region. *Glob. Biogeochem. Cycles* **23**, GB2023, <http://dx.doi.org/10.1029/2008GB003327> (2009).
- Ronov, A., Khain, V. & Balukhovskiy, S. *Atlas of Lithological–Paleogeographical Maps of the World: Mesozoic and Cenozoic of Continents and Oceans* (eds Barsukov, V. L. & Lavirov, N. P.) (Moscow Editorial Publishing Group VNIIG Zharubezh-Geologia, Moscow, 1989).
- Sloan, L. C., Walker, J. C. G., Moore, T. C. Jr & Rea, D. K. Possible methane-induced polar warming in the early Eocene. *Nature* **357**, 320–322 (1992).
- Kalkreuth, W. D. *et al.* Petrological, palynological and geochemical characteristics of Eureka Sound group coals (Stenkul Fiord, southern Ellesmere Island, Arctic Canada). *Int. J. Coal Geol.* **30**, 151–182 (1996).
- Röhl, U., Brinkhuis, H., Sluijs, A. & Fuller, M. in *The Cenozoic Southern Ocean: Tectonics, Sedimentation, and Climate Change between Australia and Antarctica* (eds Exon, N. F., Malone, M. & Kennett, J. P.) 113–125 (Geophysical Monograph Series 151, American Geophysical Union, 2004).
- Schaefer, K., Zhang, T., Bruhwiler, L. & Barrett, A. P. Amount and timing of permafrost carbon release in response to climate warming. *Tellus B* **63**, 165–180 (2011).
- Elberling, B., Christensen, H. H. & Hansen, B. U. High nitrous oxide production from thawing permafrost. *Nature Geosci.* **3**, 332–335 (2010).
- Pagani, M., Caldeira, K., Berner, R. & Beerling, D. The role of terrestrial plants in limiting atmospheric CO<sub>2</sub> decline over the past 24 million years. *Nature* **460**, 85–88 (2009).
- Clymo, R. S. The limits to peat bog growth. *Phil. Trans. R. Soc. Lond. B* **303**, 605–654 (1984).
- Zachos, J. C., Dickens, G. R. & Zeebe, R. E. An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature* **451**, 279–283 (2008).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was funded by the US National Science Foundation under award ATM-0513402/0513421 to R.M.D. and D.P., and EAR-0628358 to M.P. D.J.B. acknowledges support from a Royal Society–Wolfson Research Merit Award.

**Author Contributions** R.M.D. conceived the permafrost–hyperthermal model with input from S.G., M.P., D.T., D.P. and D.J.B. S.G. developed the cyclostratigraphic framework and performed MTM and SSA analyses. R.M.D., D.T. and D.P. designed the numerical modelling scheme and D.T. analysed the GCM results. D.J.B. and R.M.D. developed the changing GHG concentration scenarios for the model simulations. K.S. and T.Z. refined the carbon calculations. R.M.D. was the primary author and all co-authors contributed to the writing and response to reviewers.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to R.M.D. ([deconto@geo.umass.edu](mailto:deconto@geo.umass.edu)).

## METHODS

**Climate-vegetation-soil model.** The current (2011) version of the GENESIS v. 3.0 GCM<sup>32</sup> is interactively coupled to the BIOME4 equilibrium vegetation model<sup>33</sup>. The model has been tested extensively in present day and palaeoclimate scenarios and produces distributions of potential vegetation and permafrost close to observations. The atmospheric component has 18 vertical layers, a spectral resolution of T31 (~3.75°), and uses an adapted version of the NCAR CCM3 solar and thermal infrared radiation code<sup>34</sup>. The model atmosphere is coupled to 2° × 2° surface models including a 50-m slab ocean model with prognostic sea surface temperatures, diffusive heat transport, and dynamic-thermodynamic sea ice. Terrestrial land surface components include multi-layer snow and soil models, and a land-surface-transfer scheme (LSX) that calculates momentum transfer and fluxes of energy and water between the atmosphere and ice, snow, soil surfaces, and upper and lower vegetation canopies. This version of the GCM has a sensitivity to 2 × CO<sub>2</sub> of 2.9 °C, without vegetation, GHG, or ice sheet feedbacks.

In the absence of Eocene vegetation data with global coverage, potential equilibrium vegetation distributions are predicted by BIOME4. The model predicts the distribution, community structure and biogeochemistry of 27 biomes using monthly climatologies of temperature, precipitation and clouds simulated by the GCM. In turn, the simulated vegetation provides the physical land-surface attributes in the GCM. Climate-biome simulations (Table 1) were run for 40 years to allow equilibration and climatological means were calculated from the last 10 years.

The soil model<sup>35</sup> has six layers with the bottom of each layer 0.5, 0.15, 0.25, 0.75, 1.75 and 4.25 m below the surface. The model predicts evolving vertical profiles of temperature, moisture and frozen water in response to the surface climatology, snow cover, soil hydraulic properties, drainage through the bottom of the soil column, and coupling between frozen pore space and hydraulics. Given the spatial resolution of our model (2° × 2°) and unresolved details of Eocene topography, we estimate total permafrost region (Table 1) rather than the smaller (sub-grid scale) area of actual frozen ground. In our modern control simulation, this produces a permafrost distribution and area (23.3 × 10<sup>6</sup> km<sup>2</sup>) very close to observations (22.8 × 10<sup>6</sup> km<sup>2</sup>)<sup>36</sup>. In our warm Eocene simulations, perennially frozen ground is mostly limited to layers 5 and 6, with an extensive active layer above. Calculated permafrost areas are based on layer 6.

**Model boundary conditions and inputs.** Model boundary conditions include a new hybrid reconstruction of early Eocene global geography, with Antarctic shorelines and elevations in a prior reconstruction<sup>37</sup> replaced with the new (larger) Antarctic reconstruction<sup>21</sup> (Supplementary Information). A suite of simulations (Table 1) was run with progressively increasing levels of GHGs, assuming background levels increased steadily from the Palaeocene until the Early Eocene Climatic Optimum (~51 Myr ago), with punctuated, transient increases at the hyperthermals<sup>8,11,38</sup>. Prescribed CO<sub>2</sub>, CH<sub>4</sub> and N<sub>2</sub>O atmospheric mixing ratios are based on a combination of proxy estimates of CO<sub>2</sub> and estimates of CH<sub>4</sub> and N<sub>2</sub>O simulated using a three-dimensional Earth system model incorporating chemistry-climate feedbacks<sup>8</sup>. Excluding the hyperthermals, background levels of late Palaeocene-early Eocene atmospheric CO<sub>2</sub> are assumed to range from 500 p.p.m.v. to 1,000 p.p.m.v. (refs 11, 38), with atmospheric CH<sub>4</sub> increasing from 1,200 p.p.b.v. to 3,000 p.p.b.v. and N<sub>2</sub>O increasing from 350 p.p.b.v. to 400 p.p.b.v. in response to an enhanced hydrological cycle and expanded wetlands on a mostly ice-free world<sup>8</sup>. For comparative purposes, the combined radiative forcing potential of prescribed CO<sub>2</sub>, CH<sub>4</sub> and N<sub>2</sub>O are converted to equivalent CO<sub>2</sub> volume mixing ratios (e.p.p.m.v.) using standard expressions<sup>39,40</sup>.

Prescribed orbital configurations (Table 1) use values for eccentricity and obliquity chosen to broadly represent those occurring during high-eccentricity and high-obliquity orbital nodes. The orbital parameters do not attempt to represent specific time slices, which are uncertain in orbital solutions of Eocene age<sup>18</sup>. **Permafrost-soil organic carbon calculations.** Lacking observations of Palaeogene permafrost distribution, carbon content and topography, we estimate frozen permafrost soil carbon (PFSC) using current observations, assuming past permafrost conditions are similar to today, differing only in the total permafrost area. We divide frozen PFSC into peatland, near surface and deep pools such that:

$$\text{PFSC} = A_{\text{tot}} R \rho D \quad (1)$$

where  $A_{\text{tot}}$  is the total permafrost area from each simulation,  $R$  is the ratio of pool area to total permafrost area,  $\rho$  is the permafrost carbon density, and  $D$  is the thickness of the PFSC deposit. Peatland  $R$  is the ratio of modern wetland area to total permafrost area<sup>22</sup> and near surface  $R$  is one minus peatland  $R$ . An analysis of the reconstructed topography (Supplementary Information) indicates that ~15–30% of the permafrost land area has surface slopes (<0.1%) suitable for the development of peatland, consistent with modern wetland area fraction for permafrost regions. The value for peatland  $R$  is conservative, considering that

estimates of Palaeocene-Eocene wetland area are up to 3 times greater than today<sup>8,23,24,41</sup>. Deep PFSC deposits form under special conditions of especially fast sedimentation: deltaic deposits and the aeolian deposition that formed the modern yedoma deposits.  $R$  for deep deposits is the ratio of areas with known deep deposits to total permafrost area<sup>22</sup>. The near surface and deep deposits have a  $\rho$  of either 21 or 30 kg m<sup>-3</sup> based on typical observed values<sup>22,42–44</sup> and peatland  $\rho$  is based on observed peat densities<sup>45</sup>.  $D$  for near-surface PFSC deposits is either the thickness of the lowest layer of the model (2.5 m) or the average simulated permafrost thickness,  $D_{\text{ave}}$ .  $D_{\text{ave}}$  is the maximum depth of the soil column (4.25 m) minus the average active layer thickness defined by the zero degree isotherm.  $D_{\text{ave}}$  ranges between 2.4 and 3.1 m with an average of 2.6 m. For the modern control run, we accounted for the much shorter accumulation times by setting a maximum PFSC depth of 3 m, resulting in a  $D$  of 1.4 m.  $D$  for deep deposits is 22 m (ref. 22).

The peatland  $D$  assumes mature, steady state peat development. A peat bog consists of upper acrotelm and lower catotelm layers, divided by the depth of the summer water table<sup>30</sup>. The frozen PFSC is located in the catotelm below the water table. Peatland  $D$  is the thickness of the catotelm, which is a balance of organic matter settling out of the acrotelm and loss of matter due to decay<sup>30</sup>:

$$\frac{\partial D}{\partial t} = \frac{p_c}{\rho} - \frac{D}{\tau} \quad (2)$$

where  $p_c$  is the flux of carbon from the acrotelm (kg m<sup>-2</sup> yr<sup>-1</sup>), and  $\tau$  is the catotelm decay rate (yr). After ~50,000 years, input balances decay and  $D$  approaches a constant, steady state value<sup>30</sup>. Continuous deposition on unglaciated terrain before the PETM and between most hyperthermals lasted >10<sup>6</sup> years, so on average, peatlands were near steady state. Setting the time derivative to zero and rearranging gives:

$$D = \frac{p_c \tau}{\rho} \quad (3)$$

We assume a  $\tau$  of 1,500 years and a  $p_c$  of 0.27 ± 0.19 kg m<sup>-2</sup> yr<sup>-1</sup> based on typical observed values<sup>30</sup> to estimate a peatland  $D$  of 2.9 ± 2.0 m.

Values of  $R$ ,  $\rho$  and  $D$  used for peatland, near-surface, and deep PFSC deposits (Supplementary Information) provide an ensemble of 12 members. The ensemble mean is our best estimate of frozen PFSC and the standard deviation is uncertainty. Subtracting PFSC from the pre-PETM simulation (experiment 2) gives the potential release for each simulation. The effect of PFSC release on atmospheric CO<sub>2</sub>, ignoring other feedbacks, assumes a relationship of 0.4602 p.p.m.v. per Pg C (ref. 27). A rate of carbon release up to 1.5 Pg C yr<sup>-1</sup> (Supplementary Information) is derived by scaling to model calculations of future permafrost<sup>27</sup> and modern observations<sup>46</sup>. Carbon isotopic values are expressed as  $[R(\text{sample})/R(\text{standard}) - 1] \times 1,000$ , where  $R$  is the <sup>13</sup>C/<sup>12</sup>C ratio of sample relative to the Pee Dee Belemnite standard.

32. Thompson, S. L. & Pollard, D. Greenland and Antarctic mass balances for present and doubled atmospheric CO<sub>2</sub> from the GENESIS Version-2 Global Climate Model. *J. Clim.* **10**, 871–900 (1997).
33. Kaplan, J. O. et al. Climate change and Arctic ecosystems: 2. Modeling, paleodata-model comparisons, and future projections. *J. Geophys. Res.* **108** (D19), 8171, <http://dx.doi.org/10.1029/2002JD002559> (2003).
34. Kiehl, J. T. et al. The National Center for Atmospheric Research Community Climate Model: CCM3. *J. Clim.* **11**, 1131–1149 (1998).
35. Thompson, S. L. & Pollard, D. A global climate model (GENESIS) with a land-surface-transfer scheme (LSX). Part I: present-day climate. *J. Clim.* **8**, 732–761 (1995).
36. Zhang, T., Heginbottom, J. A., Barry, R. G. & Brown, J. Further statistics on the distribution of permafrost and ground ice in the northern hemisphere. *Polar Geogr.* **24**, 126–131 (2000).
37. Sewall, J. O., Sloan, L. C., Huber, M. & Wing, S. Climate sensitivity to changes in land surface characteristics. *Glob. Planet. Change* **26**, 445–465 (2000).
38. Pagani, M., Caldeira, K., Archer, D. E. & Zachos, J. An ancient carbon mystery. *Science* **314**, 1556–1557 (2006).
39. Ramaswamy, V. et al. in *Climate Change 2001: The Scientific Basis* (eds Haughton, J. T. et al.) 351–416 (Cambridge Univ. Press, 2001).
40. Shi, G. Radiative forcing and greenhouse effect due to the atmospheric trace gases. *Sci. China B* **35**, 217–229 (1992).
41. Beerling, D., Berner, R. A., Mackenzie, F. T., Harfoot, M. B. & Pyle, J. A. Methane and the CH<sub>4</sub>-related greenhouse effect over the past 400 million years. *Am. J. Sci.* **309**, 97–113 (2009).
42. Dutta, K., Schuur, E. A. G., Neff, J. C. & Zimov, S. A. Potential carbon release from permafrost soils of Northeastern Siberia. *Glob. Change Biol.* **12**, 2336–2351 (2006).
43. Khvorostyanov, D. V., Ciais, P., Krinner, G. & Zimov, S. A. Vulnerability of east Siberia's frozen carbon stores to future warming. *Geophys. Res. Lett.* **35**, L10703, <http://dx.doi.org/10.1029/2008GL033639> (2008).
44. Zech, M. et al. Characterization and palaeoclimate of a loess-like permafrost palaeosol sequence in NE Siberia. *Geoderma* **143**, 281–295 (2008).
45. Price, J. S., Cagampan, J. & Kellner, E. Assessment of peat compressibility: is there an easy way? *Hydrol. Process.* **19**, 3469–3475 (2005).
46. Schuur, E. A. G. et al. The effect of permafrost thaw on old carbon release and net carbon exchange from tundra. *Nature* **459**, 556–559 (2009).



# A gigantic feathered dinosaur from the Lower Cretaceous of China

Xing Xu<sup>1,2</sup>, Kebai Wang<sup>3</sup>, Ke Zhang<sup>4</sup>, Qingyu Ma<sup>1</sup>, Lida Xing<sup>5</sup>, Corwin Sullivan<sup>1</sup>, Dongyu Hu<sup>2</sup>, Shuqing Cheng<sup>3</sup> & Shuo Wang<sup>1,6</sup>

Numerous feathered dinosaur specimens have recently been recovered from the Middle–Upper Jurassic and Lower Cretaceous deposits of northeastern China, but most of them represent small animals<sup>1</sup>. Here we report the discovery of a gigantic new basal tyrannosauroid, *Yutyrannus huali* gen. et sp. nov., based on three nearly complete skeletons representing two distinct ontogenetic stages from the Lower Cretaceous Yixian Formation of Liaoning Province, China. *Y. huali* shares some features, particularly of the cranium, with derived tyrannosauroids<sup>2,3</sup>, but is similar to other basal tyrannosauroids<sup>4–12</sup> in possessing a three-fingered manus and a typical theropod pes. Morphometric analysis suggests that *Y. huali* differed from tyrannosauroids in its growth strategy<sup>13,14</sup>. Most significantly, *Y. huali* bears long filamentous feathers, thus providing direct evidence for the presence of extensively feathered gigantic dinosaurs and offering new insights into early feather evolution.

The Tyrannosauroidae was one of the longest-lived theropod subgroups, with a fossil record extending from the Middle Jurassic to the uppermost Cretaceous<sup>2</sup>. Basal tyrannosauroids are relatively small, and gigantic ones (adult body mass greater than 1,000 kg) are almost entirely restricted to the latest Cretaceous<sup>2,15</sup>. Four tyrannosauroid taxa have recently been reported from the Lower Cretaceous of China<sup>6,8,9,15,16</sup>, although the provenance of one of them, *Raptorex kriegsteini*, has been seriously questioned<sup>17</sup>. These taxa range from 1.4 m to about 10 m in body length<sup>6,8</sup> and show considerable morphological disparity: some taxa closely resemble the highly specialized Tyrannosauridae<sup>9,16</sup>, whereas others are more similar to generalized coelurosaurs<sup>6,8</sup>. Combined with discoveries from outside China, these morphologically and taxonomically diverse basal tyrannosauroid specimens document the occurrence of a significant radiation in the early history of the group.

Here we report the discovery of a new feathered tyrannosauroid (Figs 1 and 2 and Supplementary Figs 1–3) from the Lower Cretaceous of China that is close to some Late Cretaceous tyrannosauroids in adult size (Supplementary Information). Phylogenetic analyses using two different theropod matrices place this taxon among basal tyrannosauroids, but relatively close to the Tyrannosauridae (Fig. 3 and Supplementary Information). In combination with other recent discoveries, such as that of the similarly sized *Sinotyrannus* from the Lower Cretaceous of Liaoning<sup>8</sup>, the new find demonstrates that tyrannosauroids were the dominant large predators in the middle Early Cretaceous ecosystems of northeastern China, suggesting that the ecological dominance of the group was achieved early in their evolution in some geographical regions at least.

Tyrannosauroidae Marsh, 1881

Coelurosauria sensu Gauthier, 1986

Tyrannosauroidae Osborn, 1905

*Yutyrannus huali* gen. et sp. nov.

**Etymology.** The generic name is derived from ‘yu’ (Mandarin for ‘feathers’) + ‘tyrannus’ (Latin for ‘king’ or ‘tyrant’). The specific name

‘huali’ means ‘beautiful’ in Mandarin, referring to the beauty of the plumage of this animal.

**Holotype.** ZCDM (Zhucheng Dinosaur Museum, Shandong) V5000, a semi-articulated, nearly complete skeleton. A cast of the specimen is housed at the Institute of Vertebrate Paleontology and Paleoanthropology as IVPP FV1960.

**Paratypes.** ZCDM V5001, a nearly complete, articulated skeleton; and ELDM (Erlianhaote Dinosaur Museum, Inner Mongolia) V1001, an articulated skeleton missing the tail. Casts of these specimens are housed at the Institute of Vertebrate Paleontology and Paleoanthropology as IVPP FV1961 and IVPP FV1962, respectively.

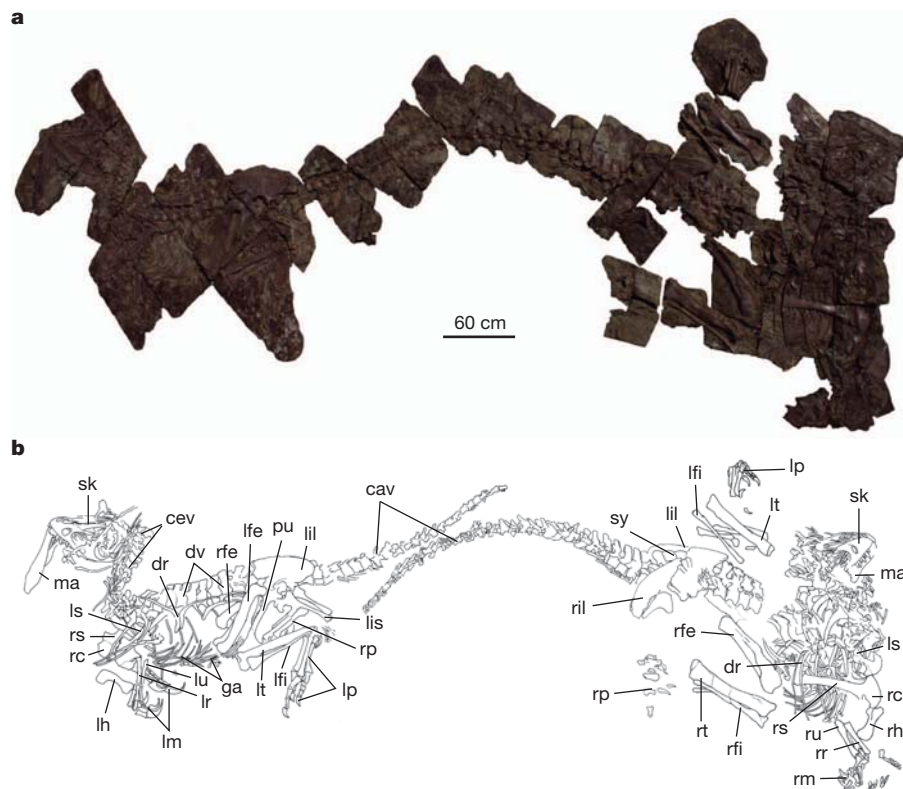
**Horizon and locality.** Batuyingzi, Beipiao, Liaoning Province, China; Lower Cretaceous Yixian Formation<sup>18</sup>.

**Diagnosis.** A gigantic tyrannosauroid distinguishable from other tyrannosauroids by the unique presence of a rugose, highly fenestrated midline crest formed by the premaxillae and nasals, an anteroventrally projecting orbital process in the area of the junction between the frontal and jugal processes of the postorbital, a large concavity on the lateral surface of the main body of the postorbital, and an external mandibular fenestra located mostly within the surangular. Also differs from *Sinotyrannus* in that the morphologically lateral surface of the maxillary process of the premaxilla faces dorsally, the maxilla lacks an anterior ramus, the maxillary fenestra is posteriorly positioned, the antorbital fossa has a posteroventrally sloping ventral margin, and the ilium has a straight dorsal margin and a postacetabular process whose ventral margin bears a lobe-like flange.

**Description and comparisons.** ZCDM V5000 probably represents an adult individual, given that the neurocentral sutures on all of the visible vertebrae are closed and the sacra are fused together. With a femoral length of 85 cm, ZCDM V5000 even exceeds the adult sizes of some Late Cretaceous tyrannosauroids, such as *Dryptosaurus* (77 cm) and *Appalachiosaurus* (79 cm). ZCDM V5000 is estimated to have had a mass of about 1,414 kg as a living animal, on the basis of an empirical equation<sup>19</sup>, and ZCDM V5001 and ELDM V1001 are estimated to have had respective masses of 596 kg and 493 kg. Both ZCDM V5001 and ELDM V1001 display fusion features, such as visible neurocentral sutures on all of the presacral vertebrae, suggesting an ontogenetic stage considerably earlier than that inferred for ZCDM V5000. On the basis of data on the growth of other large tyrannosauroids<sup>13</sup>, ELDM V1001 is estimated to be at least 8 years younger than ZCDM V5000.

The most striking cranial feature of *Y. huali* is a highly pneumatic midline crest resembling those of *Guanlong*<sup>7</sup> and the carcharodontosaurian *Concavenator*<sup>20</sup>, although in *Y. huali* the crest is formed by premaxillary and nasal portions that are only loosely articulated with each other. The dorsal margin of the crest bears a series of low prominences that are likely to be homologous to the rugosities seen in all Late Cretaceous tyrannosauroids<sup>9</sup>. The cranium of *Y. huali* also exhibits some features that occur consistently in basal, but not derived,

<sup>1</sup>Institute of Vertebrate Paleontology and Paleoanthropology, Key Laboratory of Evolutionary Systematics of Vertebrates, Chinese Academy of Sciences, 142 Xiwai Street, Beijing 100044, China. <sup>2</sup>Liaoning Paleontological Museum, Shenyang Normal University, 253 North Huanghe Street, Shenyang 110034, China. <sup>3</sup>Zhucheng Dinosaur Museum, Zhucheng, Shandong, 262200, China. <sup>4</sup>School of the Earth Sciences and Resources, China University of Geosciences, 28 Xueyuan Road, Beijing 100083, China. <sup>5</sup>Department of Biological Sciences, University of Alberta, 11145 Saskatchewan Drive, Edmonton, Alberta T6G 2E9, Canada. <sup>6</sup>Graduate University of Chinese Academy of Sciences, 19 Yuquan Road, Beijing 100049, China.



**Figure 1** | *Yutyrannus huali* (ZCDM V5000 and ZCDM V5001).

**a**, Photograph of the slab preserving ZCDM V5000 and ZCDM V5001. **b**, Line drawing of the slab. Abbreviations: cav, caudal vertebra; cev, cervical vertebra; dr, dorsal rib; dv, dorsal vertebra; ga, gastralia; lfe, left femur; lfi, left fibula; lh, left humerus; lil, left ilium; lis, left ischium; lm, left manus; lp, left pes; lr, left

radius; ls, left scapula; lt, left tibiotarsus; lu, left ulna; ma, mandible; pu, pubis; rc, right coracoid; rfe, right femur; rh, right humerus; ril, right ilium; rm, right manus; rp, right pes; rr, right radius; rs, right scapula; rt, right tibiotarsus; ru, right ulna; sk, skull; sy, synsacrum.

tyrannosauroids<sup>2,3,15</sup>. The elliptical external naris is large and posteriorly positioned; a sharp groove runs along the anterior margin of the premaxilla; the maxilla bears a sharp groove paralleling the ventral rim of the antorbital fossa; the jugal has a raised, anteroposteriorly orientated rim; and the surangular has a long anterior flange. Previous studies have suggested that some of these features are synapomorphies of the basal clade Proceratosauridae<sup>4,15</sup>, but our phylogenetic analysis optimizes several of these features as synapomorphies of the Tyrannosauroidea.

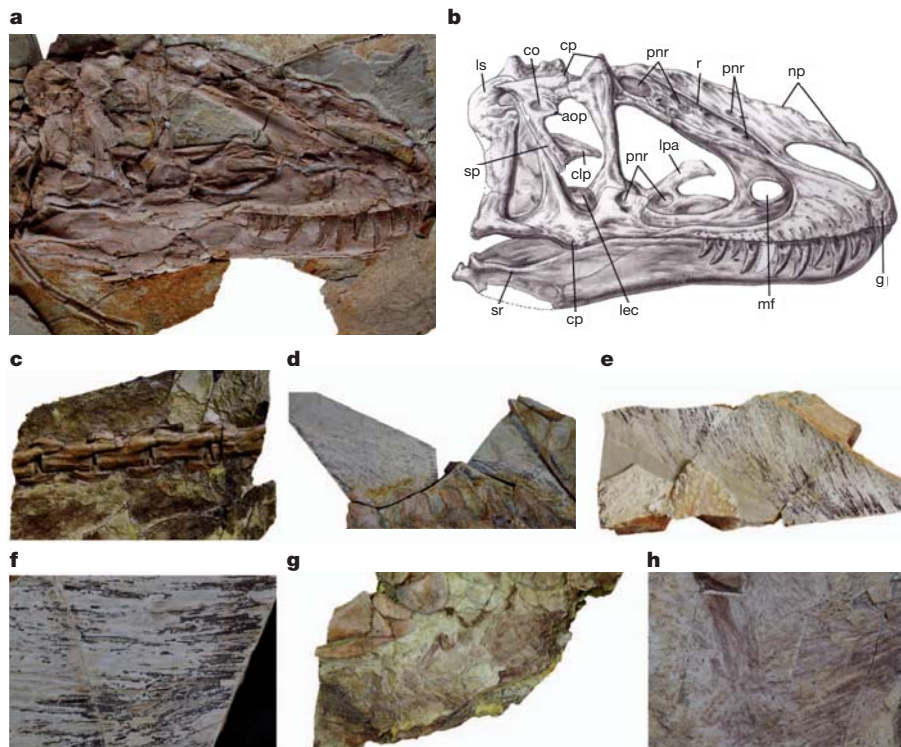
However, many other cranial features of *Y. huali* are more similar to those of derived tyrannosauroids<sup>2,3,21,22</sup>. For example, the skull is large and deep; the premaxilla has a proportionally deep main body and a maxillary process whose morphologically lateral surface faces dorsally in adults; the maxilla has a markedly convex ventral margin and a posteriorly tapering main body; the lacrimal is in the shape of a '7'; the cornual process of the lacrimal is a large conical structure; the postorbital has a wide jugal process and a suborbital process extending into the orbit; the squamosal has an anteroposteriorly orientated quadratojugal process that intrudes into the infratemporal fenestra; the quadratojugal has a large posterior process that overlaps the posterior surface of the quadrate; the external mandibular fenestra is small; the dentary has a strongly concave dorsal margin and a posteriorly located inflection point between the anterior and ventral margins in lateral view; and the surangular has a prominent horizontal ridge.

The vertebrae are not pneumatized to the degree seen in the Tyrannosauridae, but they show initial development of some features that are characteristic of derived tyrannosauroids<sup>2,21</sup>. These features include prominent flanges for ligament attachment on the anterior and posterior margins of the neural spines of the cervical and dorsal vertebrae, tall neural spines on the posterior cervical vertebrae, laterally placed prezygapophyses on the middle cervical vertebrae, and anteroposteriorly shortened dorsal vertebrae with posteriorly placed neural spines.

The shoulder girdle is in general plesiomorphic, as indicated by the relative robustness of the scapular blade, the weakly expanded distal end of the scapula, and the large coracoid foramen. The forelimbs are also similar to those of basal tyrannosauroids in retaining a typical basal coelurosaurian design, including a three-fingered manus<sup>15</sup>.

The pelvis displays several derived features<sup>15</sup>: the dorsal margin of the ilium is mostly straight, the ventral margin of the postacetabular process of the ilium bears a prominent lobe-like flange, the pubic boot is large and forms a distinct anterior expansion, and the ischium is much more slender than the pubis. The hindlimbs are generally similar to those of basal tyrannosauroids, and the distal segments are proportionally short, more similar to allosauroids and basal tyrannosauroids than to tyrannosauroids<sup>2,14</sup>.

Filamentous integumentary structures are preserved in all three specimens. Those preserved in ZCDM V5000 are evidently associated with the posterior caudal vertebrae. As preserved, they are parallel to each other and form an angle of about 30° with the long axis of the tail. The filaments are at least 15 cm long. They are too densely packed for it to be possible to determine whether they are elongate broad filamentous feathers (EBFFs) like those seen in the therizinosauroid *Beipiaosaurus*, slender monofilaments, or compound filamentous structures. Those of ZCDM V5001 are near the pelvis and pes. They are filamentous structures, but morphological details are not preserved. In ELDM V1001, integumentary filaments are visible extending from the dorsal side of the neck, and near a limb bone that is tentatively identified as a humerus. Those extending from the neck measure more than 20 cm, and those along the humerus at least 16 cm. Although feather preservation is patchy in these specimens, as occurs even in some fossil birds from the Jehol Group that undoubtedly had plumage covering most of the body, the distribution of the preserved filamentous feathers in the three specimens of *Y. huali* implies that this taxon had an extensively



**Figure 2 | Selected elements of *Y. huali* (ZCDM V5000, ZCDM V5001 and ELDM V1001).** **a**, Photograph of the skull and mandible of ELDM V1001. **b**, Line drawing of the skull and mandible of ELDM V1001. **c–h**, filamentous integumentary structures preserved in the three specimens: **c**, along the posterior caudal vertebrae of ZCDM V5000; **d**, along the cervical vertebrae of ELDM V1001; **e**, along a limb bone of ELDM V1001; **g**, **h**, near the pes of

ZCDM V5001 (**f** and **h** are close-up views). Abbreviations: aop, accessory orbital process; clp, cultriform process; co, concauity; cp, cornual process; g, groove; lec, left ectopterygoid; lpa, left palatine; ls, left squamosal; mf, maxillary fenestra; np, nasal prominences; pnr, pneumatic recesses; r, ridge; sp, suborbital process; sr, surangular ridge.

feathered integument in life. This has also been inferred for *Dilong*, and some other non-avian feathered dinosaurs<sup>1</sup>.

**Morphological variations.** Some morphological differences between ZCDM V5000, ZCDM V5001 and ELDM V1001 may represent ontogenetic variations. With increasing maturity, for example, the skull becomes deeper and more robust, the premaxilla becomes narrower and taller, the anterior portion of the premaxilla becomes more medially orientated, the lateral surface of the maxillary process of the premaxilla rotates to face dorsally, and the maxillary fenestra becomes more anteriorly located. Several other morphological variations, such as the presence of a relatively straight dorsal margin of the ilium in ZCDM V5000 and ELDM V1001 and a convex one in ZCDM V5001, are more difficult to interpret in ontogenetic terms. They may reflect individual genetic variation or sexual dimorphism.

A morphometric analysis suggests that *Y. huali* differed in its growth pattern from the highly specialized tyrannosaurids (Supplementary Information). Using femur length as a standard proxy for overall size, the scapula and ilium display negative allometry in *Y. huali* (in contrast to positive allometry and near isometry, respectively, in tyrannosaurids<sup>14</sup>). The radius, metacarpus and distal hindlimb segments are negatively allometric in both *Y. huali* and the Tyrannosauridae, but the negative allometry of the metacarpus, tibia and metatarsus is much stronger in *Y. huali* than in tyrannosaurids<sup>14</sup>.

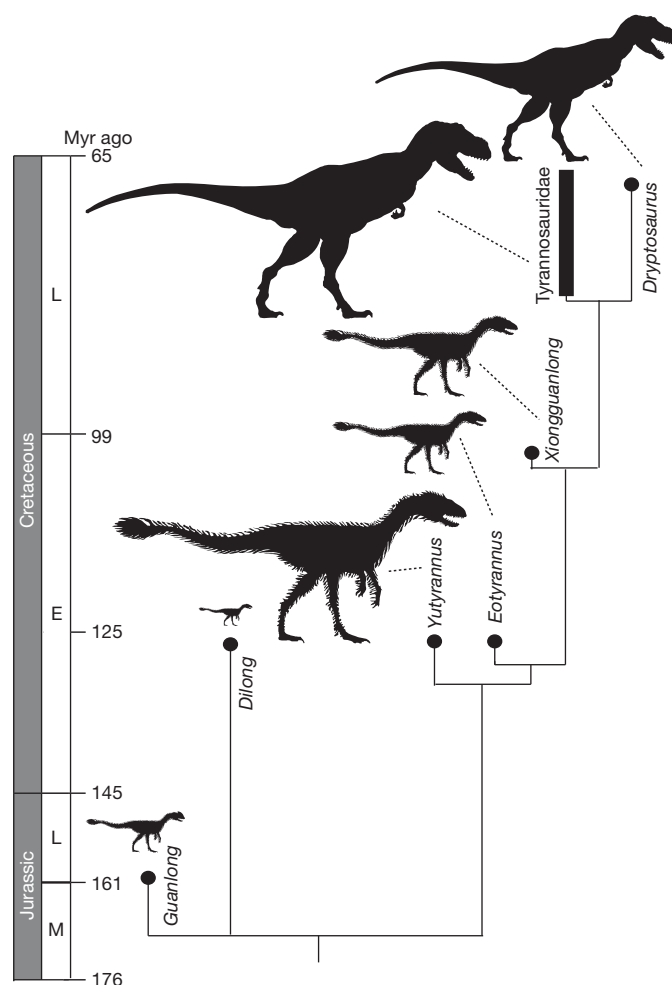
This discovery has implications for early feather evolution. Although some gigantic dinosaurs are likely to have been feathered animals<sup>23</sup>, the largest previously known non-avian dinosaur in which direct evidence for a feathery covering is available is *Beipiaosaurus*<sup>24</sup> (adult body mass about 1/40 that of ZCDM V5000). The discovery of *Y. huali* provides solid evidence for the existence of gigantic feathered dinosaurs and, more significantly, of a gigantic species with an extensive feathery covering.

Gigantism affects many aspects of animal structure and function. Extensive filamentous integumentary coverings such as feathers and

hair are partly or even primarily insulative in function, but some large mammals have become almost entirely hairless because their low surface-to-volume ratios permit them to retain metabolic heat even without a pelage (although large mammals living in cold environments, such as the bovid *Bison bison*, retain substantial fur). Gigantic tyrannosaurids have been suggested to lack an extensive feathery covering for analogous reasons<sup>6</sup>. This interpretation derives some support from reported impressions of small patches of scaly skin<sup>25,26</sup>, and there is certainly no direct fossil evidence for the presence of feathers in gigantic Late Cretaceous tyrannosaurids. The discovery of *Y. huali*, however, indicates that at least one gigantic dinosaur had an extensive insulative coat of feathers, showing in turn that drastic reduction of the plumage was not an inevitable consequence of very large body size. If Late Cretaceous tyrannosaurids such as *Tyrannosaurus rex* were similar to *Y. huali* in this respect, both basal and derived tyrannosaurid dinosaurs would differ from mammals in lacking a tendency to lose their integumentary covering as result of gigantism.

Alternatively, if scales were indeed the dominant integumentary structures in most Late Cretaceous tyrannosaurids, the presence of long feathers in the gigantic *Y. huali* could represent an adaptation to an unusually cold environment. *Y. huali* lived during a period (the Barremian–early Albian) that has been interpreted as considerably colder than the rest of the Cretaceous (a mean annual air temperature of about 10 °C in western Liaoning, in contrast with about 18 °C at a similar latitude in the Late Cretaceous)<sup>27</sup>. Most gigantic Late Cretaceous tyrannosaurids, by contrast, lived in a warm climate that was conducive to the loss of an extensive insulative feathery covering, although populations inhabiting cold environments such as the land that is now Alaska would have been a notable exception<sup>28,29</sup>. It is possible that the extent and nature of the integumentary covering changed over time in response to shifts in body size and the temperature of the environment throughout tyrannosaurid evolutionary history, as has clearly occurred in some mammalian taxa<sup>30</sup>. However, it must be noted that





**Figure 3 | A simplified cladogram showing the systematic position of *Y. huali* among the Tyrannosauroidea.** Silhouettes indicate body size and possible extent of plumage. Different tyrannosauroids seem to have attained gigantic body size independently in the Early and Late Cretaceous, but only in the Early Cretaceous is there direct evidence of a gigantic form with an extensively feathered integument. This may reflect the relatively cold climate of the middle Early Cretaceous. See also Supplementary Information.

the plumage is only partly preserved in all three known specimens of *Y. huali*, and the possibility that the feathers had only a restricted distribution on the body cannot be completely excluded. If this was so, the feathers might have functioned primarily as display structures as in some other non-avian theropod groups<sup>1</sup>.

Received: 26 November 2011; accepted 27 January 2012.

- Xu, X. & Guo, Y. The origin and early evolution of feathers: insights from recent paleontological and neontological data. *Vert. Palasiatica* **47**, 311–329 (2009).
- Holtz, T. R. in *The Dinosauria* 2nd edn (eds Weishampel, D.B., Dodson, P. & Osmólska, H.) 111–136 (Univ. California Press, Berkeley, 2004).
- Currie, P. J. Cranial anatomy of tyrannosaurid dinosaurs from the Late Cretaceous of Alberta, Canada. *Acta Palaeontol. Pol.* **48**, 191–226 (2003).
- Rauhut, O. W. M., Milner, A. C. & Moore-Fay, S. Cranial osteology and phylogenetic position of the theropod dinosaur *Proceratosaurus bradleyi* (Woodward, 1910) from the Middle Jurassic of England. *Zool. J. Linn. Soc.* **158**, 155–195 (2010).
- Rauhut, O. W. M. A tyrannosauroid dinosaur from the Upper Jurassic of Portugal. *Palaeontology* **46**, 903–910 (2003).
- Xu, X. *et al.* Basal tyrannosauroids from China and evidence for protofeathers in tyrannosauroids. *Nature* **431**, 680–684 (2004).

- Xu, X. *et al.* A basal tyrannosauroid dinosaur from the Late Jurassic of China. *Nature* **439**, 715–718 (2006).
- Ji, Q., Ji, S. A. & Zhang, L. J. First known large tyrannosauroid theropod from the Early Cretaceous Jehol Biota in northeastern China. *Geol. Bull. China* **28**, 1369–1374 (2009).
- Li, D. Q., Norell, M. A., Gao, K.-Q., Smith, N. D. & Makovicky, P. J. A longirostrine tyrannosauroid from the Early Cretaceous of China. *Proc. R. Soc. Lond. B* **277**, 183–190 (2010).
- Averianov, A. O., Krasnolutskii, S. A. & Ivantsov, S. V. A new basal coelurosaur (Dinosauria: Theropoda) from the Middle Jurassic of Siberia. *Proc. Zool. Inst. RAS* **314**, 42–57 (2010).
- Benson, R. B. J. New information on *Stokesosaurus*, a tyrannosauroid (Dinosauria: Theropoda) from North America and the United Kingdom. *J. Vertebr. Paleontol.* **28**, 732–750 (2008).
- Hutt, S., Naish, D., Martill, D. M., Barker, M. J. & Newbery, P. A preliminary account of a new tyrannosauroid theropod from the Wessex Formation (Early Cretaceous) of southern England. *Cretac. Res.* **22**, 227–242 (2001).
- Erickson, G. M. *et al.* Gigantism and comparative life-history parameters of tyrannosaurid dinosaurs. *Nature* **430**, 772–775 (2004).
- Currie, P. J. Allometric growth in tyrannosaurids (Dinosauria: Theropoda) from the Upper Cretaceous of North America and Asia. *Can. J. Earth Sci.* **40**, 651–665 (2003).
- Brusatte, S. *et al.* Tyrannosaur paleobiology: new research on ancient exemplar organisms. *Science* **329**, 1481–1485 (2010).
- Sereno, P. C. *et al.* Tyrannosaurid skeletal design first evolved at small body size. *Science* **326**, 418–422 (2009).
- Fowler, D. W., Woodward, H. N., Freedman, E. A., Larson, P. L. & Horner, J. R. Reanalysis of ‘*Raptorex kriegsteini*’: a juvenile tyrannosaurid dinosaur from Mongolia. *PLoS ONE* **6**, e21376 (2011).
- Swisher, C. C. *et al.* Further Support for a Cretaceous age for the feathered-dinosaur beds of Liaoning, China: New 40Ar/39Ar dating of the Yixian and Tuchengzi Formations. *Chin. Sci. Bull.* **47**, 135–138 (2002).
- Christiansen, P. & Fariña, R. A. Mass prediction in theropod dinosaurs. *Hist. Biol.* **16**, 85–92 (2004).
- Ortega, F., Escaso, F. & Sanz, J. L. A bizarre, humped Carcharodontosauria (Theropoda) from the Lower Cretaceous of Spain. *Nature* **467**, 203–206 (2010).
- Brochu, C. A. Osteology of *Tyrannosaurus rex*: insights from a nearly complete skeleton and high-resolution computed tomographic analysis of the skull. *J. Vertebr. Paleontol. Mem.* **7**, 1–138 (2003).
- Brusatte, S. L., Carr, T. D., Erickson, B. R., Bever, G. S. & Norell, M. A. A long-snouted, multihorned tyrannosaurid from the late Cretaceous of Mongolia. *Proc. Natl Acad. Sci. USA* **106**, 17261–17266 (2009).
- Xu, X., Tang, Q.-W., Wang, J.-M., Zhao, X.-J. & Tan, L. A gigantic bird-like dinosaur from the late Cretaceous of China. *Nature* **447**, 844–847 (2007).
- Xu, X., Tang, Z.-L. & Wang, X.-L. A therizinosauroid dinosaur with integumentary structures from China. *Nature* **399**, 350–354 (1999).
- Paul, G. S. in *Tyrannosaurus rex, the Tyrant King* (eds Carpenter, K. & Larson, P.E.) 354–368 (Indiana Univ. Press, 2008).
- Currie, P. J., Badamgarav, D. & Koppelhus, E. B. The first Late Cretaceous footprints from the Nemegt Locality in the Gobi of Mongolia. *Ichnos* **10**, 1–13 (2003).
- Amiot, R. *et al.* Oxygen isotopes of east Asian dinosaurs reveal exceptionally cold Early Cretaceous Climates. *Proc. Natl Acad. Sci. USA* **108**, 5179–5183 (2011).
- Fiorillo, A. R. & Gangloff, R. A. Theropod teeth from the Prince Creek Formation (Cretaceous) of northern Alaska, with speculations on Arctic dinosaur paleoecology. *J. Vertebr. Paleontol.* **20**, 675–682 (2000).
- Spicer, R. A. & Herman, A. B. The Late Cretaceous environment of the Arctic: a quantitative reassessment based on plant fossils. *Paleogeogr. Palaeoclimatol. Palaeoecol.* **295**, 423–442 (2010).
- Deng, T. *et al.* Out of Tibet: Pliocene woolly rhino suggests high-plateau origin of ice age megaherbivores. *Science* **333**, 1285–1288 (2011).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank L. Zhang for discussions, R. Li, H. Zang and X. Ding for illustrations, and H. Wang, L. Xiang and R. Cao for preparing the specimens. We thank the Zhucheng Municipal Government and Erlanhaote Municipal Government for support. This study was supported by grants from the National Natural Science Foundation of China and Special Funds For Major State Basic Research Projects of China.

**Author Contributions** X.X. designed the project. X.X., K.W., K.Z., Q.M., L.X., C.S., D.H., S.C. and S.W. performed the research. X.X., C.S. and Q.M. wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to X.X. ([xingxu@vip.sina.com](mailto:xingxu@vip.sina.com)).

# A universal model for mobility and migration patterns

Filippo Simini<sup>1,2,3</sup>, Marta C. González<sup>4</sup>, Amos Maritan<sup>2</sup> & Albert-László Barabási<sup>1,5,6</sup>

Introduced in its contemporary form in 1946 (ref. 1), but with roots that go back to the eighteenth century<sup>2</sup>, the gravity law<sup>1,3,4</sup> is the prevailing framework with which to predict population movement<sup>3,5,6</sup>, cargo shipping volume<sup>7</sup> and inter-city phone calls<sup>8,9</sup>, as well as bilateral trade flows between nations<sup>10</sup>. Despite its widespread use, it relies on adjustable parameters that vary from region to region and suffers from known analytic inconsistencies. Here we introduce a stochastic process capturing local mobility decisions that helps us analytically derive commuting and mobility fluxes that require as input only information on the population distribution. The resulting radiation model predicts mobility patterns in good agreement with mobility and transport patterns observed in a wide range of phenomena, from long-term migration patterns to communication volume between different regions. Given its parameter-free nature, the model can be applied in areas where we lack previous mobility measurements, significantly improving the predictive accuracy of most of the phenomena affected by mobility and transport processes<sup>11–23</sup>.

In analogy with Newton's law of gravity, the gravity law assumes that the number of individuals  $T_{ij}$  that move between locations  $i$  and  $j$  per unit time is proportional to some power of the population of the source ( $m_i$ ) and destination ( $n_j$ ) locations, and decays with the distance  $r_{ij}$  between them as

$$T_{ij} = \frac{m_i^\alpha n_j^\beta}{f(r_{ij})} \quad (1)$$

where  $\alpha$  and  $\beta$  are adjustable exponents and the deterrence function  $f(r_{ij})$  is chosen to fit the empirical data. Occasionally  $T_{ij}$  is interpreted as the probability rate of individuals travelling from  $i$  to  $j$ , or an effective coupling between the two locations<sup>24</sup>. Despite its widespread use, the gravity law has notable limitations:

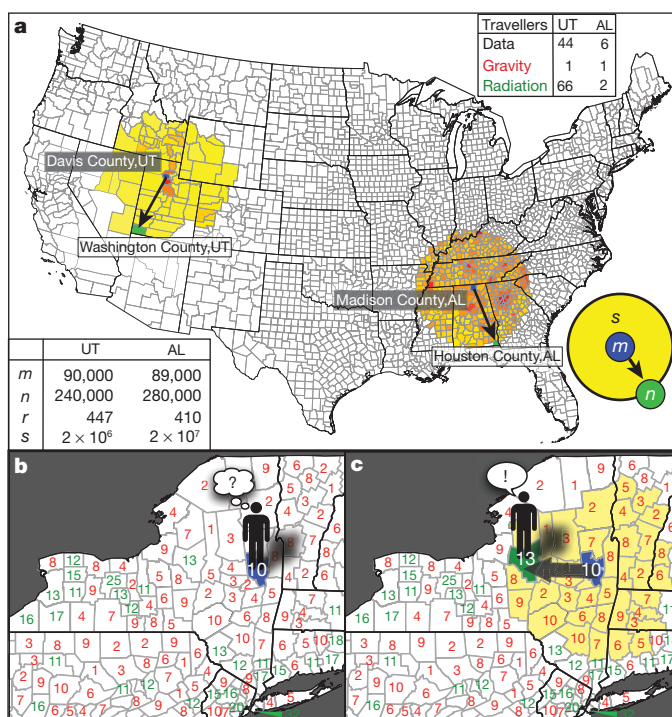
Limitation one, we lack a rigorous derivation of (1). Whereas entropy maximization<sup>25</sup> leads to (1) with  $\alpha = \beta = 1$ , it fails to offer the functional form of  $f(r)$ .

Limitation two, lacking theoretical guidance, practitioners use a range of deterrence functions (power law or exponential) and up to nine parameters to fit the empirical data<sup>5,7,8,11,14</sup>.

Limitation three, as (1) requires previous traffic data to fit the parameters  $[\alpha, \beta, \dots]$ , it is unable to predict mobility in regions where we lack systematic traffic data, areas of major interest in modelling of infectious diseases.

Limitation four, the gravity law has systematic predictive discrepancies. Indeed, in Fig. 1a we highlight two pairs of counties with similar origin and destination populations and comparable distance, so according to (1) the flux between them should be the same. Yet, the US census (see Supplementary Information) documents an order of magnitude difference between the two fluxes: only 6 individuals commute between the two Alabama counties, whereas 44 do in Utah.

Limitation five, equation (1) predicts that the number of commuters increases without limit as we increase the destination population  $n_j$ , yet



**Figure 1 | The radiation model.** **a**, To demonstrate the limitations of the gravity law we highlight two pairs of counties, one in Utah (UT) and the other in Alabama (AL), with similar origin ( $m$ , blue) and destination ( $n$ , green) populations and comparable distance  $r$  between them (see bottom left table). The gravity law predictions were obtained by fitting equation (1) to the full commuting data set, recovering the parameters  $[\alpha, \beta, \gamma] = [0.30, 0.64, 3.05]$  for  $r < 119$  km, and  $[0.24, 0.14, 0.29]$  for  $r > 119$  km of ref. 14. The fluxes predicted by (1) are the same because the two county pairs have similar  $m$ ,  $n$  and  $r$  (top right table). Yet the US census 2000 reports a flux that is an order of magnitude greater between the Utah counties, a difference correctly captured by the radiation model (**b**, **c**). **b**, The definition of the radiation model: an individual (for example, living in Saratoga County, New York) applies for jobs in all counties and collects potential employment offers. The number of job opportunities in each county ( $j$ ) is  $n_j/n_{\text{jobs}}$ , chosen to be proportional to the resident population  $n_j$ . Each offer's attractiveness (benefit) is represented by a random variable with distribution  $p(z)$ , the numbers placed in each county representing the best offer among the  $n_j/n_{\text{jobs}}$  trials in that area. Each county is marked in green (red) if its best offer is better (lower) than the best offer in the home county (here  $z = 10$ ). **c**, An individual accepts the closest job that offers better benefits than his home county. In the shown configuration the individual will commute to Oneida County, New York, the closest county whose benefit  $z = 13$  exceeds the home county benefit  $z = 10$ . This process is repeated for each potential commuter, choosing new benefit variables  $z$  in each case.

<sup>1</sup>Center for Complex Network Research and Department of Physics, Biology and Computer Science, Northeastern University, Boston, Massachusetts 02115, USA. <sup>2</sup>Dipartimento di Fisica "G. Galilei", Università di Padova, CNISM and INFN, via Marzolo 8, 35131 Padova, Italy. <sup>3</sup>Institute of Physics, Budapest University of Technology and Economics, Budafoki út 8, Budapest, H-1111, Hungary. <sup>4</sup>MIT, Department of Civil and Environmental Engineering, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA. <sup>5</sup>Center for Cancer Systems Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. <sup>6</sup>Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA.

the number of commuters cannot exceed the source population  $m_i$ , highlighting the gravity law's analytical inconsistency (see Supplementary Information, Section 4).

Limitation six, being deterministic, the gravity law cannot account for fluctuations in the number of travellers between two locations.

Motivated by these known limitations, alternative approaches like the intervening opportunity model<sup>26</sup> or the random utility model<sup>27</sup> (Supplementary Information, Section 7) have been proposed. Although derived from first principles, these models continue to contain context-specific tunable parameters, and their predictive power is at best comparable to the gravity law<sup>28</sup>.

Here we introduce a modelling framework that relies on first principles and overcomes the problems of limitations one to six of the gravity law. Whereas commuting is a daily process, its source and destination is determined by job selection, a decision made over longer timescales. Using the natural partition of a country into counties (for which commuting data are collected), we assume that job selection consists of two steps (Fig. 1 b, c).

Step one, an individual seeks job offers from all counties, including his/her home county. The number of employment opportunities in each county is proportional to the resident population,  $n$ , assuming that there is one job opening for every  $n_{\text{jobs}}$  individuals. We capture the benefits of a potential employment opportunity with a single number,  $z$ , randomly chosen from distribution  $p(z)$  where  $z$  represents a combination of income, working hours, conditions, etc. Thus, each county with population  $n$  is assigned  $n/n_{\text{jobs}}$  random numbers,  $z_1, z_2, \dots, z_{\lfloor n/n_{\text{jobs}} \rfloor}$ , accounting for the fact that the larger a county's population, the more employment opportunities it offers.

Step two, the individual chooses the closest job to his/her home, whose benefits  $z$  are higher than the best offer available in his/her home county. Thus lack of commuting has priority over the benefits, that is, individuals are willing to accept lesser jobs closer to their home.

This process, applied in proportion to the resident population in each county, assigns work locations to each potential commuter, which in turn determines the daily commuting fluxes across the country. The model has three unknown parameters: the benefit distribution  $p(z)$ , the job density  $n_{\text{jobs}}$ , and the total number of commuters,  $N_c$ . We show, however, that the commuting fluxes  $T_{ij}$  are independent of  $p(z)$  and  $n_{\text{jobs}}$ , and the remaining free parameter,  $N_c$ , does not affect the flux distribution, making the model parameter-free. As the model can be formulated in terms of radiation and absorption processes (see Supplementary Information, Section 2), we will refer to it as the radiation model. To analytically predict the commuting fluxes we consider locations  $i$  and  $j$  with population  $m_i$  and  $n_j$  respectively, at distance  $r_{ij}$  from each other, and we denote with  $s_{ij}$  the total population in the circle of radius  $r_{ij}$  centred at  $i$  (excluding the source and destination population). The average flux  $T_{ij}$  from  $i$  to  $j$ , as predicted by the radiation model (see Supplementary Information, Section 2), is

$$\langle T_{ij} \rangle = T_i \frac{m_i n_j}{(m_i + s_{ij})(m_i + n_j + s_{ij})} \quad (2)$$

which is independent of both  $p(z)$  and  $n_{\text{jobs}}$ . Hence (2) represents the fundamental equation of the radiation model, the proposed alternative to the gravity law (1). Here  $T_i \equiv \sum_{j \neq i} T_{ij}$  is the total number of commuters

that start their journey from location  $i$ , which is proportional to the population of the source location, hence  $T_i = m_i(N_c/N)$ , where  $N_c$  is the total number of commuters and  $N$  is the total population in the country (Fig. 2g).

Equation (2) resolves limitations one to six of the gravity law: it has a rigorous derivation (resolving limitation one) and has no free parameters (bypassing limitations two and three). To understand the origin of limitation four, we note that a key difference between the radiation model (2) and the gravity law (1) is that the variable of (2) is not the distance  $r_{ij}$  but  $s_{ij}$ . Thus the commuting flux depends not only on  $m_i$  and  $n_j$  but also on the population  $s_{ij}$  of the region surrounding the

source location. For uniform population density  $s_{ij} \approx m_i r_{ij}^2$  and  $n = m$ , (2) reduces to the gravity law (1) with  $f(r) = r^2$ ,  $\gamma = 4$  and  $\alpha + \beta = 1$ . The non-uniform population density, however, is key to resolving the problem of limitation four: equation (2) predicts an order of magnitude difference in Alabama and Utah, in line with the census data (see Fig. 1a). Indeed the population density around Utah is significantly lower than the United States average, thus work opportunities within the same radius are ten times smaller in Utah than in Alabama, implying that commuters in Utah have to travel farther to find comparable employment opportunities. Note also that equation (2) predicts that the number of travellers leaving from a location with population  $m$  to one with  $n \rightarrow \infty$  saturates at  $T_{n \rightarrow \infty} = \frac{m^2}{(m+s)} + O(\frac{1}{n}) \leq m$ , resolving the unphysical divergence highlighted in limitation five. Finally,  $T_{ij}$  in the radiation model is a stochastic variable, predicting not only the average flux between two locations (2), but also its variance (see Supplementary Information, Section 2), resolving the problem of limitation six.

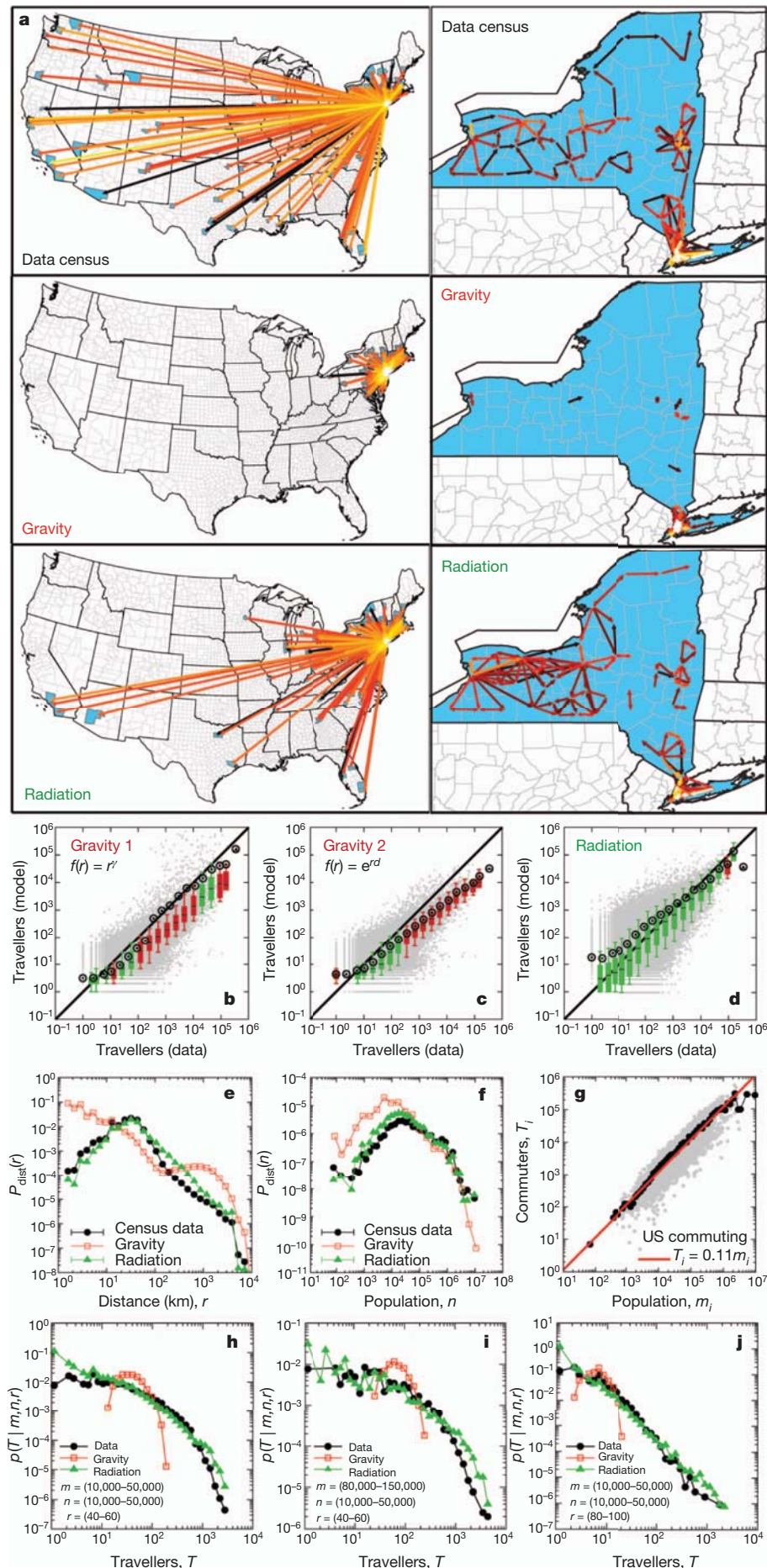
To explore the radiation model's ability to predict the correct commuting patterns, in Fig. 2a we show the commuting fluxes with more than ten travellers originating from New York County. The destinations predicted by the gravity law<sup>14</sup> are all within 400 km from the origin, missing all long distance and many medium distance trips. The gravity law's local performance is equally poor: within the State of New York it grossly overestimates fluxes in the vicinity of New York City and underestimates the fluxes in the rest of the state (Fig. 2a, right column). The radiation model offers a more realistic approximation to the observed commuting patterns, both nationally and state-wide (Fig. 2a, bottom panels). To quantify the observed differences, we compare the measured and the predicted non-zero commuting fluxes for all pairs of counties in the United States. We find that both standard implementations of the gravity law<sup>11,14</sup> ( $f(r) = r^2$  and  $f(r) = e^{dr}$ , where  $d$  is a fitting parameter with the unit of an inverse length needed to ensure a dimensionless argument to the exponential function) significantly underestimate the high flux commuting patterns, often by an order of magnitude or more (Fig. 2b, c). In contrast, the average fluxes predicted by the radiation model are within the error bars despite the observed six orders of magnitude span in commuting fluxes (Fig. 2d).

The systematic failure of the gravity law is particularly evident if we measure the probability  $P_{\text{dist}}(r)$  of a trip between locations at distance  $r$  (Fig. 2e), and the probability of trips towards a destination with population  $n$ ,  $P_{\text{dest}}(n)$  (Fig. 2f). For  $P_{\text{dist}}(r)$  the radiation model clearly follows the peak around 40 km in the census data. The prediction based on the gravity law lacks this peak and thus it overestimates by three orders of magnitude the number of short distance trips. Similarly, the gravity law overestimates the low  $n$  values of  $P_{\text{dest}}(n)$  by nearly an order of magnitude.

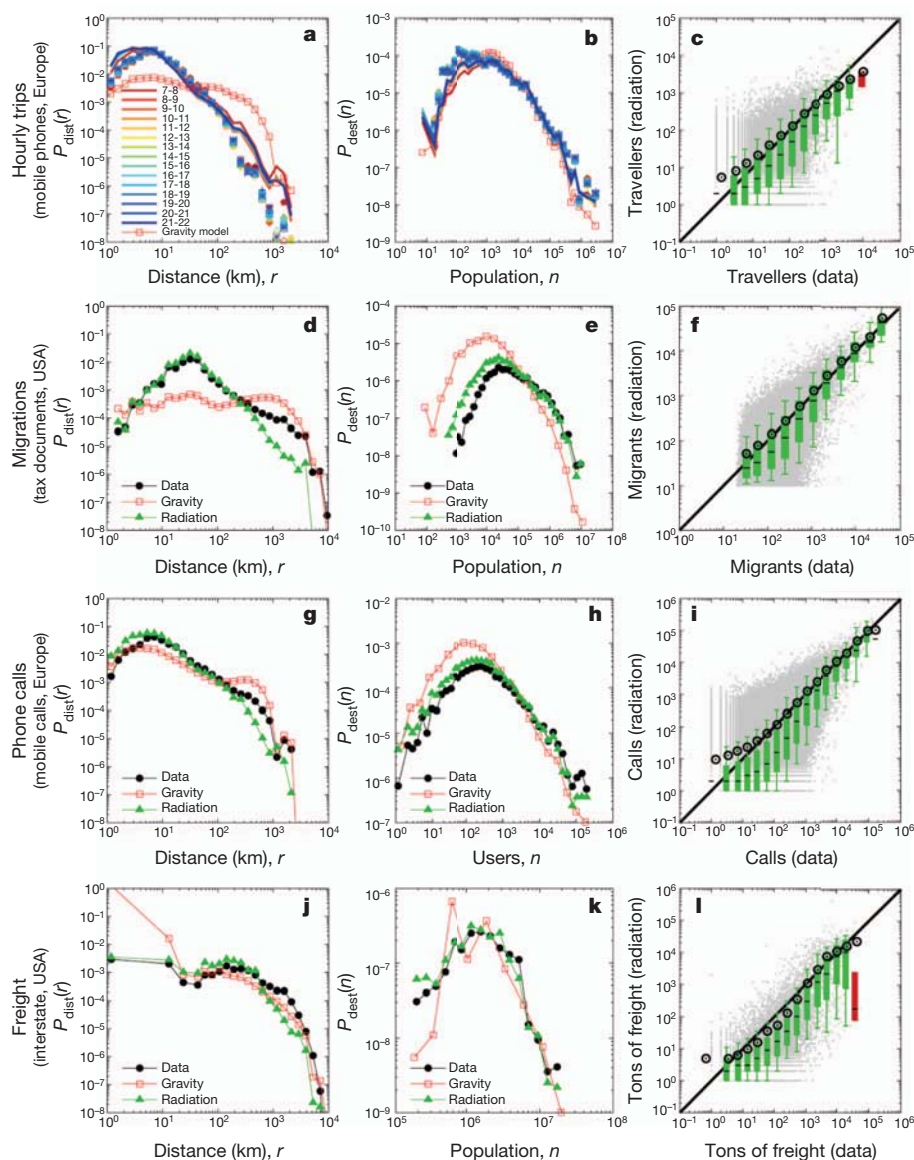
Another important mobility measure is the conditional probability  $p(T|m, n, r)$  to observe a flux of  $T$  individuals from a location with population  $m$  to a location with population  $n$  at a distance  $r$ . The gravity law predicts a highly peaked  $p(T|m, n, r)$  distribution around the average  $\langle T \rangle_{mnr} = \sum_T p(T|m, n, r)T$  (Fig. 2h–j), because, according to (1) pairs of locations with the same  $(m, n, r)$  have the same flux. In contrast the radiation model predicts a broad  $p(T|m, n, r)$  distribution, in reasonable agreement with the data.

To show the generality of the model in Fig. 3 we test its performance for four socio-economic phenomena: hourly travel patterns, migrations, communication patterns and commodity flows. We find that the radiation model offers an accurate quantitative description of mobility and transport spanning a wide range of time scales (hourly mobility, daily commuting, yearly migrations), capturing diverse processes (commuting, intra-day mobility, call patterns, trade), collected via a wide range of tools (census, mobile phones, tax documents) on different continents (America, Europe). The agreement with data of such diverse nature is somewhat surprising, suggesting that the hypotheses behind the model capture fundamental decision mechanisms that, directly or indirectly, are relevant to a wide span of mobility and transport-driven processes.





**Figure 2 | Comparing the predictions of the radiation model and the gravity law.** **a**, National mobility fluxes with more than ten travellers originating from New York County (left panels) and the high intensity fluxes (over 1,100 travellers) within the state of New York (right panels). Arrows represent commuters fluxes, the colour capturing flux intensity: black, 10 individuals (fluxes below ten travellers are not shown for clarity), white, >10,000 individuals. The top panels display the fluxes reported in US census 2000, the central panels display the fluxes fitted by the gravity law with  $f(r) = r^{-2}$ , and the bottom panels display the fluxes predicted by the radiation model. **b–d**, Comparing the measured flux,  $T_{ij}^{\text{data}}$ , with the predicted flux,  $T_{ij}^{\text{GM}}$  and  $T_{ij}^{\text{Rad}}$ , for each pair of counties. We compare the census data with two formulations of the gravity law,  $f(r) = e^{-dr}$  (**c**) and  $f(r) = r^{-2}$  (**b**), and with the radiation model (**d**). Grey points are scatter plot for each pair of counties. A box is coloured green if the line  $y = x$  lies between the 9th and the 91st percentiles in that bin and is red otherwise. The black circles correspond to the mean number of predicted travellers in that bin. **e**, Probability of a trip between two counties that are at distance  $r$  (in km) from each other,  $P_{\text{dist}}(r)$ . **f**, Probability of a trip towards a county with population  $n$ ,  $P_{\text{dest}}(n)$ . **g**, The number of commuters in a county,  $T_i$ , is proportional to its population,  $m_i$ . **h–j**, Conditional probability  $p(T | m, n, r)$  to observe a flow of  $T$  individuals from a location with population  $m$  to a location with population  $n$  at a distance  $r$  for three triplets  $(m, n, r)$ . The gravity law predicts a highly peaked distribution around the average value  $\langle T \rangle_{mnr}$ , in disagreement with census data and the radiation model, which both show a broad distribution.



**Figure 3 | Beyond commuting.** **a–c**, Testing the radiation model on hourly trips extracted from a mobile phone database of a western European country. The anonymized billing records<sup>29,30</sup> cover the activity of approximately ten million subscribers. We analysed a 6-month period, recording the user locations with tower resolution hourly between 7am and 10pm, identifying all trips between municipalities. **a**, Probability of a trip between two municipalities at distance  $r$ ,  $P_{\text{dist}}(r)$ , shown for 14-hourly time intervals. Radiation model predictions are solid lines; gravity model's aggregated fit over 24 h is a red line with empty squares. **b**, Probability of a trip towards a municipality with population  $n$ ,  $P_{\text{dest}}(n)$ . **c**, Comparing the measured flux,  $T_{ij}^{\text{data}}$ , with the predicted flux,  $T_{ij}^{\text{Rad}}$ , for each pair of municipalities with  $T_{ij}^{\text{data}}, T_{ij}^{\text{Rad}} > 0$ , for commuting trips extracted by identifying each user's home and workplace from the locations

where the user made the most calls. **d–f**, Testing (2) on long-term migration patterns, capturing the number of individuals that relocated from one US county to another during tax years 2007–2008 as reported by the US Internal Revenue Service. **g–i**, Phone call volume between municipalities extracted from the anonymized mobile phone database. The number of phone calls between users living in different municipalities during a period of 4 weeks resulted in 38,649,153 calls placed by 4,336,217 users. We aggregated the data to obtain the total number of calls between every pair of municipalities. **j–l**, Commodity flows in the US extracted from the Freight Analysis Framework (FAF), which offers a comprehensive picture of freight movement among US states and major metropolitan areas by all modes of transportation. For each data set we measured the quantities discussed in **a–c**.

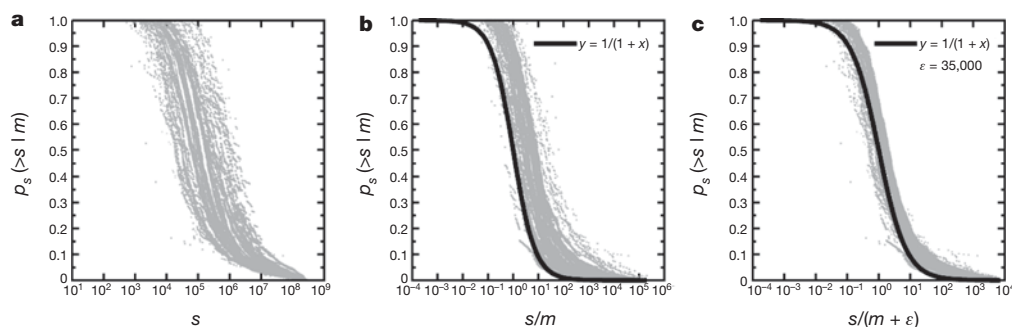
To illustrate the effect of the heterogeneous population distribution on commuting fluxes, in Supplementary Fig. 8a–f we show the commuting landscape generated by (2) from the perspective of two individuals, one in Davis county, Utah, and the other in Clayton County, Georgia, with comparable populations of 238,994 and 236,517, respectively. If the population was uniformly distributed, the landscape seen by a potential employee would be simple: the farther is a job, the less desirable it is (Supplementary Fig. 8a, d). Yet, the observed variations in population density significantly alter the local commuting landscape, as shown in Supplementary Fig. 8b, e where we coloured the US counties based on their distance to the commuter's home county (Supplementary Fig. 8a, d) and then moved

them closer or further from the origin so that the new distance reflects the true likelihood of representing a commuting destination.

Despite the observed differences in the perspective of individual commuters, the radiation model helps us uncover a previously unsuspected scale-invariance in commuting patterns. Indeed, according to (2), the probability of one trip from  $i$  to  $j$  (equal to  $T_{ij}/T_i$ ) is scale invariant under the transformation  $m_i \rightarrow \lambda m_i$ ,  $n_j \rightarrow \lambda n_j$  and  $s_{ij} \rightarrow \lambda s_{ij}$ . Empirical evidence for this statistical self-similarity is offered in Fig. 4a, b (see also Supplementary Information, Section 8).

In summary, the superior performance of the radiation model can significantly improve the accuracy of predictive tools in all areas affected by mobility and transport processes<sup>11,12</sup>, from epidemiology<sup>13</sup> and





**Figure 4 | Unveiling the hidden self-similarity in human mobility.** **a**, The probability  $p_s(>s|m)$  to observe a trip from a location with population  $m$  to a destination in the region beyond a population  $s$  from the origin ( $m$  varies between 200 and 2,000,000). **b**, According to the radiation model

$p_s(>s|m) = 1/(1+s/m)$ , a homogeneous function of the ratio  $s/m$ . Plotting  $p_s(>s|m)$  versus  $s/m$ , the curves approach the theoretical result  $y = 1/(1+x)$ . **c**, The collapse improves if we account for the home field advantage in job search by always adding  $\epsilon = 35,000$  to the population of the commuter's home county.

spreading processes<sup>17</sup> to urban geography<sup>18–21</sup> and flow of resources in economics<sup>22</sup>. The parameter-free modelling platform we introduced can predict commuting and transport patterns even in areas where such data are not collected systematically, as it relies only on population densities, which is relatively accurately estimated throughout the globe.

Despite its superior performance the radiation model can absorb further improvements. For example, consider the fact that an individual has a home-field advantage when searching for jobs in the home county, being more familiar with local employment opportunities. We can incorporate this by adding  $\epsilon/n_{\text{jobs}}$  additional employment opportunities to his/her home county, achieved through an effective increase  $m \rightarrow m + \epsilon$  of the home county population, so the adjusted law is now invariant under the  $(m + \epsilon) \rightarrow \lambda(m + \epsilon)$  and  $s \rightarrow \lambda s$  transformation. We find that the rescaling of the commuting probability improves dramatically (Fig. 4c), indicating that the home field advantage offers an effective boost in employment opportunities that is equivalent with an additional  $\epsilon = 35,000$  individuals in the home county population. Furthermore, the adjusted radiation model shows a better or equally good agreement with the real data in all tested measures (Supplementary Fig. 6), demonstrating that equation (2) is not a rigid end point of our approach, but offers a platform that can be improved upon in specific environments.

Received 8 August 2011; accepted 13 January 2012.

Published online 26 February 2012.

- Zipf, G. K. The  $P_1P_2/D$  hypothesis: on the intercity movement of persons. *Am. Sociol. Rev.* **11**, 677–686 (1946).
- Monge, G. *Mémoire sur la Théorie des Déblais et de Remblais. Histoire de l'Académie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la même année 666–704* (De l'Imprimerie Royale, 1781).
- Barthélemy, M. Spatial networks. *Phys. Rep.* **499**, 1–101 (2010).
- Erlander, S. & Stewart, N. F. *The Gravity Model in Transportation Analysis: Theory and Extensions* (VSP, 1990).
- Jung, W. S., Wang, F. & Stanley, H. E. Gravity model in the Korean highway. *EPL* **81**, 48005 (2008).
- Thiemann, C., Theis, F., Grady, D., Brune, R. & Brockmann, D. The structure of borders in a small world. *PLoS ONE* **5**, e15422 (2010).
- Kaluza, P., Kölsch, A., Gastner, M. T. & Blasius, B. The complex network of global cargo ship movements. *J. R. Soc. Interf.* **7**, 1093–1103 (2010).
- Krings, G., Calabrese, F., Ratti, C. & Blondel, V. D. Urban gravity: a model for intercity telecommunication flows. *J. Stat. Mech.* **2009**, L07003 (2009).
- Expert, P., Evans, T. S., Blondel, V. D. & Lambiotte, R. Uncovering space-independent communities in spatial networks. *Proc. Natl Acad. Sci. USA* **108**, 7663–7668 (2011).
- Pöyhönen, P. A tentative model for the volume of trade between countries. *Weltwirtschaftliches Arch.* **90**, 93–100 (1963).
- Balcan, D. et al. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl Acad. Sci. USA* **106**, 21484–21489 (2009).
- Helbing, D. Traffic and related self-driven many-particle systems. *Rev. Mod. Phys.* **73**, 1067–1141 (2001).

- Colizza, V., Barrat, A., Barthélemy, M. & Vespignani, A. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proc. Natl Acad. Sci. USA* **103**, 2015–2020 (2006).
- Viboud, C. et al. Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science* **312**, 447–451 (2006).
- Ferguson, N. M. et al. Strategies for mitigating an influenza pandemic. *Nature* **442**, 448–452 (2006).
- Xu, X. J., Zhang, X. & Mendes, J. F. F. Impacts of preference and geography on epidemic spreading. *Phys. Rev. E* **76**, 056109 (2007).
- Lind, P. G., Da Silva, L. R., Andrade, J. S. Jr & Herrmann, H. J. Spreading gossip in social networks. *Phys. Rev. E* **76**, 036117 (2007).
- Roth, C., Kang, S. M., Batty, M. & Barthélemy, M. Structure of urban movements: polycentric activity and entangled hierarchical flows. *PLoS ONE* **6**, e15923 (2011).
- Makse, H. A., Havlin, S. & Stanley, H. E. Modelling urban growth patterns. *Nature* **377**, 608–612 (1995).
- Bettencourt, L. M. A., Lobo, J., Helbing, D., Kühnert, C. & West, G. B. Growth, innovation, scaling, and the pace of life in cities. *Proc. Natl Acad. Sci. USA* **104**, 7301–7306 (2007).
- Batty, M. The size, scale, and shape of cities. *Science* **319**, 769–771 (2008).
- Garlaschelli, D., Di Matteo, T., Aste, T., Caldarelli, G. & Loffredo, M. I. Interplay between topology and dynamics in the World Trade Web. *The Eur. Phys. J. B* **57**, 159–164 (2007).
- Eubank, S. et al. Modelling disease outbreaks in realistic urban social networks. *Nature* **429**, 180–184 (2004).
- Krueckeberg, D. A. & Silvers, A. L. *Urban Planning Analysis: Methods and Models* (Wiley, 1974).
- Wilson, A. G. The use of entropy maximising models in the theory of trip distribution, mode split and route split. *J. Transp. Econ. Policy* 108–126 (1969).
- Stouffer, S. A. Intervening opportunities: a theory relating mobility and distance. *Am. Sociol. Rev.* **5**, 845–867 (1940).
- Block, H. D. & Marschak, J. *Random Orderings and Stochastic Theories of Responses* (Cowles Foundation, 1960).
- Rogerson, P. A. Parameter estimation in the intervening opportunities model. *Geogr. Anal.* **18**, 357–360 (1986).
- González, M. C., Hidalgo, C. A. & Barabási, A. L. Understanding individual human mobility patterns. *Nature* **453**, 779–782 (2008).
- Onnela, J. P. et al. Structure and tie strengths in mobile communication networks. *Proc. Natl Acad. Sci. USA* **104**, 7332–7336 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank J. P. Bagrow, A. Fava, F. Giannotti, Y.-R. Lin, J. Menche, Z. Neda, D. Pedreschi, D. Wang, G. Wilkerson and D. Bauer for many discussions, and N. Ferguson for prompting us to look into the gravity law. A.M. and F.S. acknowledge the Cariparo foundation for financial support. This work was supported by the Network Science Collaborative Technology Alliance sponsored by the US Army Research Laboratory under Agreement Number W911NF-09-2-0053; the Office of Naval Research under Agreement Number N000141010968; the Defense Threat Reduction Agency awards WMD BRBA07-J-2-0035 and BRBA08-Per4-C-2-0033; and the James S. McDonnell Foundation 21st Century Initiative in Studying Complex Systems.

**Author Contributions** All authors designed and did the research. F.S. analysed the empirical data and performed the numerical calculations. A.M. and F.S. developed the analytical calculations. A.-L.B. was the lead writer of the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to A.-L.B. (alb@neu.edu) and A.M. (amos.maritan@pd.infn.it).



# Endospore abundance, microbial growth and necromass turnover in deep sub-seafloor sediment

Bente Aa. Lomstein<sup>1</sup>, Alice T. Langerhuus<sup>1</sup>, Steven D'Hondt<sup>2</sup>, Bo B. Jørgensen<sup>3</sup> & Arthur J. Spivack<sup>2</sup>

Two decades of scientific ocean drilling have demonstrated widespread microbial life in deep sub-seafloor sediment, and surprisingly high microbial-cell numbers. Despite the ubiquity of life in the deep biosphere, the large community sizes and the low energy fluxes in this vast buried ecosystem are not yet understood<sup>1,2</sup>. It is not known whether organisms of the deep biosphere are specifically adapted to extremely low energy fluxes or whether most of the observed cells are in a dormant, spore-like state<sup>3</sup>. Here we apply a new approach—the D,L-amino-acid model—to quantify the distributions and turnover times of living microbial biomass, endospores and microbial necromass, as well as to determine their role in the sub-seafloor carbon budget. The approach combines sensitive analyses of unique bacterial markers (muramic acid and D-amino acids) and the bacterial endospore marker, dipicolinic acid, with racemization dynamics of stereo-isomeric amino acids. Endospores are as abundant as vegetative cells and microbial activity is extremely low, leading to microbial biomass turnover times of hundreds to thousands of years. We infer from model calculations that biomass production is sustained by organic carbon deposited from the surface photosynthetic world millions of years ago and that microbial necromass is recycled over timescales of hundreds of thousands of years.

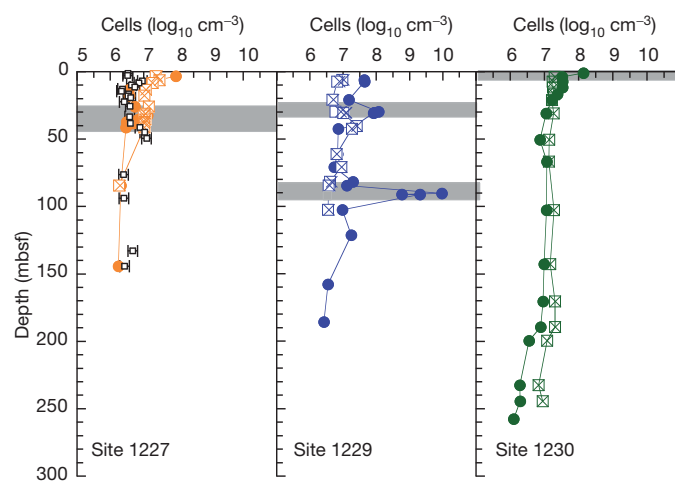
Deep sub-surface sediment material was obtained from the eastern tropical Pacific during the Ocean Drilling Program (ODP) Leg 201 expedition with DS JOIDES Resolution. Coring sites ranged from the continental shelf off the coast of Peru to ocean depths of 5,000 m. The expedition recovered sediment at depths of up to 420 metres below the sea floor (mbsf), and this sediment was found to be up to 35 million years old.

We report estimates of high bacterial endospore numbers in the order of  $10^7$  endospores per  $\text{cm}^3$  (Fig. 1) in deep sediment and sediment that is up to 10 million years old. We used two different analyses to quantify endospores: muramic acid, which is a unique building block in cell walls of both bacteria and endospores; and dipicolinic acid, which is uniquely formed by endospores. In the first analysis, muramic acid of endospores is calculated as the total muramic acid (Supplementary Fig. 1) minus muramic acid from vegetative cells. Vegetative cells (and intact but recently dead cells) were quantified by acridine orange direct counts (AODCs)<sup>4</sup>. Endospore muramic acid is converted to endospore numbers by the use of cell-specific conversion factors that are obtained from the literature (Supplementary Information). It remains unclear whether bacteria or archaea predominate in the studied sediment<sup>5,6</sup>. We therefore use two extreme scenarios for our estimates of endospore abundance, either that bacteria completely dominate<sup>6</sup> (bacterial dominance) or that 90% are archaea<sup>5</sup> (archaeal dominance). Archaea do not contain muramic acid in the cell wall. We also assume that 35% of the bacteria are Gram positive and 65% are Gram negative, as was found throughout the sediment column at Site 1229 (ref. 7).

Muramic-acid-based endospore numbers are  $0.2 \times 10^7$  to  $3 \times 10^7$  endospores per  $\text{cm}^3$ . Archaeal or bacterial dominance did not affect the estimated endospore numbers because muramic acid levels calculated from AODC only marginally contributes to the measured

muramic acid concentrations (Fig. 1). At Site 1227 we also estimated endospore numbers from dipicolinic acid concentrations assuming an average dipicolinic-acid content of  $2.2 \times 10^{-16}$  mol spore<sup>-1</sup> (ref. 8). Dipicolinic-acid-based endospore numbers are  $0.3 \times 10^7$  to  $1.0 \times 10^7$  endospores per  $\text{cm}^3$  and confirmed the muramic-acid-based estimates by a mean deviation factor of 4.6. There was a significant positive correlation between the two endospore estimates from muramic acid and dipicolinic acid as judged from a statistical analysis of their depth trend (regression of muramic acid and dipicolinic acid;  $P = 0.0008$ ;  $R^2 = 0.7297$ ). To our knowledge, this is the first time that such comparative data are published for endospores in environmental samples.

This discovery of high endospore abundances raises the question of why endospores have not been detected previously and whether total cell abundance in the deep biosphere, including endospores, has been globally underestimated. Previous studies of the deep biosphere did not quantify endospores. Endospores are unlikely to be stained by fluorescent DNA dyes such as acridine orange<sup>9,10</sup> or by ribosomal RNA staining techniques such as catalysed reporter deposition fluorescence in situ hybridization because endospore walls are impermeable<sup>6</sup>. In this study we processed samples with 3 N HCl (for 4 h at 95 °C), which completely extracts muramic acid and dipicolinic acid from endospores<sup>11</sup>. Our results suggest that endospores are as abundant as vegetative prokaryotes in this deep marine biosphere. Studies in other oceanographic regions are needed to clarify how the earlier global estimate of  $3.5 \times 10^{30}$  cells<sup>12</sup> should be adjusted to account for this.



**Figure 1 | Profiles of AODCs and estimated endospore numbers on the Peruvian continental shelf (sites 1227 and 1229) and in the Trench in Peru (site 1230).** Filled circles, AODC; coloured open squares, muramic-acid-based estimated number of endospores (bacterial dominance); crosses, muramic-acid-based estimated number of endospores (archaeal dominance); black open squares, dipicolinic-acid-based estimated number of endospores ( $n = 6$ ). Grey shading, sulphate-methane transition zones. Error bars, s.d.

<sup>1</sup>Department of Bioscience, Section for Microbiology, Aarhus University, Building 1540, Ny Munkegade 114, DK-8000 Aarhus C, Denmark. <sup>2</sup>Graduate School of Oceanography, University of Rhode Island, 100A, Horn Building, Narragansett, Rhode Island 02882, USA. <sup>3</sup>Center for Geomicrobiology, Department of Bioscience, Aarhus University, Building 1535, Ny Munkegade 114, DK-8000 Aarhus C, Denmark.

It has been suggested that bacterial necromass production is responsible for enhanced organic matter preservation in highly productive regions such as the upwelling system off the coast of Peru<sup>13</sup>. Hence, whether microbial necromass is a major component<sup>13</sup> or an insignificant component<sup>14</sup> of sedimentary organic matter is an important question. We address this question by comparing the total measured amino acid carbon with the amino acid carbon estimated from vegetative cells and endospores, based on published cellular protein contents (Supplementary Information) and AODCs. Independent of whether bacteria or archaea dominate the microbial community, vegetative cells and endospores contribute less than 3% and 1% to amino acid carbon, respectively (Supplementary Fig. 2). As expected, microbial necromass accounts for the remaining >96% of amino acid carbon. The concentration of bulk organic matter progressively decreases with sediment depth and age. Amino acids contribute 4–5% to the total organic carbon in the youngest sediment and <0.1% in sediment that is several millions of years old (Supplementary Fig. 3), thus showing the preferential degradation of amino acids compared to bulk organic matter. Hence, the down-core decrease in amino acid concentrations is the result of a dynamic amino acid turnover in which amino acid degradation slightly exceeds amino acid formation. Our data suggest that the original amino acids that are derived from the water column are degraded soon after deposition and that sub-surface amino acids are produced *in situ* by new microbial biomass formation.

Almost all of the amino acids in living cells are present in the L-isomeric configuration. The only known exception is in bacteria, as they uniquely produce specific D-amino acids for incorporation into their cell-wall complex<sup>15</sup>. In marine sediment there is an additional source of D-amino acids derived from racemization, which is the slow stereo-chemical conversion of L-amino acids to D-amino acids and vice versa (see Supplementary Information). Bacterial cultures isolated from the deep sub-seafloor have been screened for D- and L-forms of aspartic acid. The D:L-Asp ratios are relatively constant with an average of 0.09 (s.e.m. = 0.02) (Supplementary Table 2). The half-life of abiotic chemical racemization (D:L = 0.33 is the ratio obtained when chemical racemization is half way to unity<sup>16</sup>) is 0.1–0.4 million years for Asp at the *in situ* temperature (Fig. 2a and Supplementary Information).

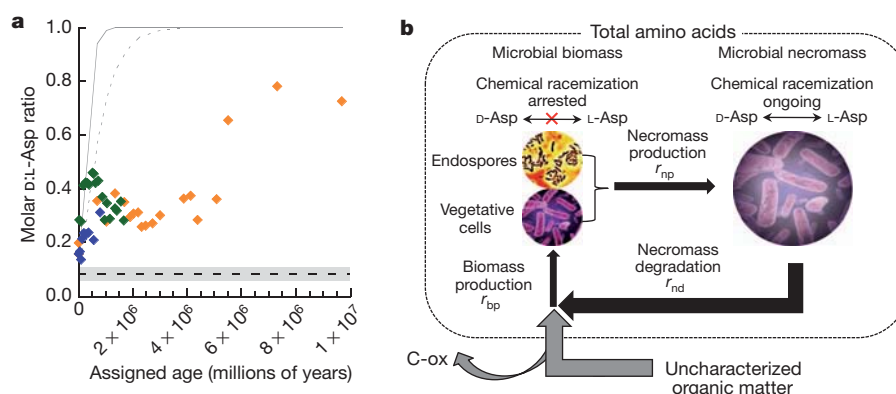
D:L-Asp ratios for sites 1227, 1229 and 1230 are higher than D:L-Asp ratios in cultures and increase in sediment that is older than 4 million years (Fig. 2a). This indicates that D-Asp is produced both biologically and by chemical racemization in sediment that is up to 10 million years old. Thus, the turnover time of Asp in the sediment is of a similar magnitude to the half-life for chemical racemization between L-Asp and D-Asp.

Constant low D:L ratios of Asp have been reported previously from vent systems<sup>17</sup> and Holocene-epoch sediments<sup>18</sup>. In those studies low D:L ratios were considered evidence of bacterial activity, but the data were not used quantitatively.

We present a D:L model from which the turnover time of amino acids in living cells and microbial necromass, as well as their role in sedimentary carbon cycling, can be quantified (Fig. 2b; see the mathematical model in Supplementary Information). Inputs to the model are the measured mean D:L-Asp ratios in cultured bacteria from the deep biosphere (Fig. 2a and Supplementary Table 2), the fraction of Asp associated with living microorganisms (Supplementary Information), and the racemization half-life of the Asp<sup>16</sup>. We use Asp because the D:L-Asp ratio in bacterial cultures varies within a limited range (Fig. 2a) and because all of our measured D:L-Asp ratios deviate from the average ratios in bacterial cultures. Such a deviation is a prerequisite for accurate model calculations of turnover time.

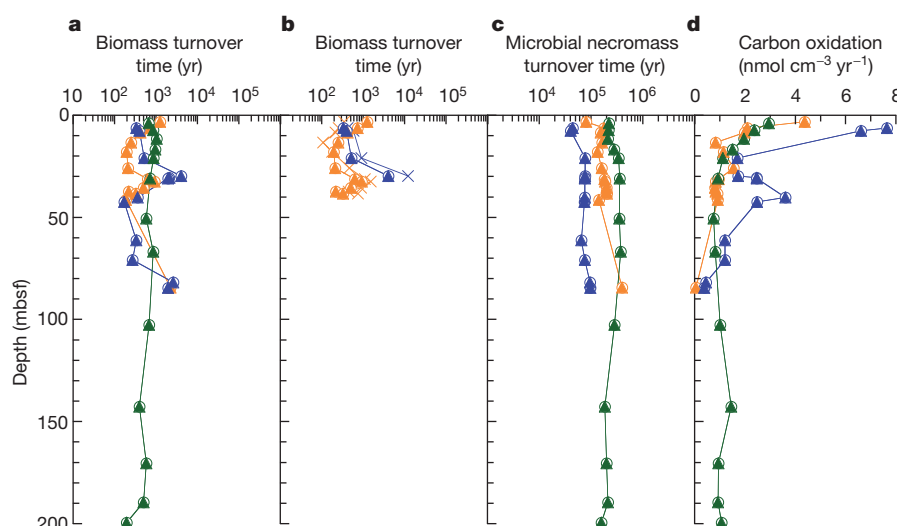
The primary output of the D:L model is the turnover time of microbial necromass, which can be used to model microbial biomass turnover time (Supplementary Information equation (25)). A quasi-steady state in microbial biomass is assumed because AODC cell numbers are relatively constant within the 5-m intervals used in the model (Fig. 1). Total carbon oxidation rates of amino acids and uncharacterized total organic carbon, and the total carbon incorporation into microbial biomass (Fig. 2b), are estimated from necromass degradation rates and an assumed growth yield (carbon incorporation efficiency). In our calculations we use an 11% growth yield, which is the average growth yield for anaerobic heterotrophic bacteria<sup>19</sup>. At steady state, when microbial biomass production is equal to necromass degradation (Supplementary equation (5)), carbon must be incorporated into vegetative cells from uncharacterized total organic carbon to maintain balance (Fig. 2b). All model estimates are performed for two extremes in microbial community composition; that is, for bacterial dominance and archaeal dominance.

The biomass turnover time,  $T_b$ , of vegetative cells is equal to cell carbon divided by the rate of cellular carbon incorporation. We use biomass-turnover time instead of generation time, because we do not know whether the cells are dividing (as indicated by the term generation time) or whether they are turning over their biomass and persist without dividing. The  $T_b$  values of vegetative cells estimated from the D:L model are less than  $0.2 \times 10^3$  to  $4.0 \times 10^3$  years, irrespective of whether bacteria or archaea dominate (Fig. 3a). These results are supported by estimates of the  $T_b$  of sulphate-reducing bacteria (SRB; less than  $0.1 \times 10^3$  to  $1.5 \times 10^3$  years) at sites 1227 and 1229 (with one exception at site 1229 at which the  $T_b$  is  $12 \times 10^3$  years; Fig. 3b). The  $T_b$  of SRB is estimated from sulphate-reduction rates modelled from



**Figure 2 | D:L-Asp model.** **a**, **b**, Turnover of amino acids in microbial biomass and necromass are quantified by application of the time constant ( $k_i$ ) for chemical racemization of Asp in necromass and the known constant D:L-Asp in bacteria. **a**, D:L-Asp versus assigned age of the sediment (Supplementary Information): orange, site 1227; blue, site 1229; green, site 1230. Mean D:L-Asp in bacterial cultures (dashed line in the shaded box; Supplementary Table 2)

and s.e.m. (shaded box;  $n = 7$ ). Curves show D:L-Asp generated by chemical racemization at the highest (solid line) and lowest (dotted line) temperatures encountered. **b**, Visualizes additional outputs of the D:L model: necromass production, necromass degradation, carbon oxidation (C-ox) and biomass production rates.



**Figure 3** | D:L-model estimates of microbial biomass turnover times, necromass turnover times and carbon-oxidation rates. **a**, Microbial-biomass turnover times. **b**, Comparison between D:L-modelled biomass turnover times, and modelled biomass turnover times of sulphate reducing bacteria (crosses).

**c**, Necromass turnover times. **d**, Carbon-oxidation rates. Orange, site 1227; blue, site 1229; green, site 1230. Open circles, bacterial dominance; filled triangles, archaeal dominance of the microbial community (the two symbols overlap).

sulphate profiles and a 13% contribution of SRB to the microbial community in the sulphate zone<sup>20</sup> (Supplementary Information). The two independent estimates of  $T_b$ , from D:L modelling and from sulphate modelling, show almost identical down-core changes at sites 1227 and 1229 (Fig. 3b), and this underscores the capability of our new D:L modelling approach. At site 1230 we have only two estimates of  $T_b$  based on sulphate reduction and they were considerably shorter than the D:L modelled  $T_b$  (data not shown).

Previous estimates of microbial generation times in deep sub-seafloor sediment are in the range of  $1 \times 10^3$  to  $2 \times 10^3$  years (ref. 12). These estimates are based on the rates of organic carbon supply and an assumed 20% growth yield. If an 11% growth yield is applied, as in our D:L model, this gives longer generation times of  $2 \times 10^3$  to  $3 \times 10^3$  years, and these generation times are still consistent with results of the present study. Irrespective of the approach used to estimate mean generation times in the deep sub-seafloor, these times are far longer than the generation times found in any other ecosystem<sup>12</sup>.

Sub-seafloor necromass is turned over on timescales that are 100-fold longer than the turnover of living biomass. Necromass turnover times are in the range of  $0.4 \times 10^5$  to  $4.0 \times 10^5$  years, irrespective of whether bacteria or archaea are assumed to dominate. This is still 10–100-fold shorter than the age of the oldest sediment (6.9 million years; Fig. 3c) and is consistent with our assumption of a quasi-steady state. It has been suggested that bacterial necromass may be an important source of recalcitrant, and hence preserved, organic matter in sediments of high-productivity regions<sup>13,21,22</sup>. This apparent contradiction between previous studies and our results is probably due to differences in timescales among studies, which range from less than a few thousand years in earlier studies<sup>13,21,22</sup> to several millions of years in the present study. Hence, preservation of sub-seafloor necromass on a shorter timescale does not preclude efficient mineralization on a longer timescale.

At all sites, D:L-Asp-modelled carbon oxidation rates decrease steeply in the upper part of the sediment (Fig. 3d). In the deeper part of site 1230 (70–200 mbsf), carbon oxidation rates remain relatively high compared to sites 1227 and 1229 (Fig. 3d). Site 1230 is located in the subduction zone and differed from the two other sites by the presence of gas hydrates at depths below 70 mbsf (ref. 23). The accumulation of methane hydrate and the more compressed sulphate-methane zonation at site 1230 compared to the other sites indicate that site 1230 has a higher carbon and energy flux in the deeper parts of the sediment.

Areal carbon oxidation rates in the sulphate zone are  $1\text{--}9 \mu\text{mol cm}^{-2} \text{yr}^{-1}$  (irrespective of bacterial or archaeal dominance) based on D:L modelling. These rates are remarkably similar to independently determined carbon oxidation rates based on sulphate profiles<sup>1</sup> ( $2\text{--}5 \mu\text{mol cm}^{-2} \text{yr}^{-1}$ ; Table 1). This supports the validity of the D:L model and gives confidence that the model is reliable also in zones below the depth of sulphate depletion.

The growth yields of the sulphate-reducing community can be estimated from biomass-production rates derived from D:L modelling, a 13% contribution of sulphate-reducing bacteria to the microbial community<sup>20</sup> and the carbon flux (carbon oxidation rate plus carbon incorporation into microbial biomass) to the sulphate-reducing bacteria, modelled from sulphate profiles (Supplementary equation (27)). The resultant average growth yield for sulphate-reducing bacteria is  $13 \pm 8\%$ , which is in agreement with the average growth yield of 11% derived from data in the literature and applied to the entire bacterial community. At the exceedingly low metabolic rates found in the present study (Fig. 3) it is unknown whether this growth yield leads to net growth and cell division or instead to turnover of the microbial-cell mass without division.

It is generally assumed that microbial communities in deep sub-surface sediments primarily rely on energy from the photosynthetic

**Table 1** | Estimated carbon oxidation rates at ODP Leg 201 sites

ODP Leg 201 site	Water depth (m)	Relevant depth interval for sulphate reduction (mbsf)	Potential C-ox by net sulphate reduction <sup>1</sup> ( $\mu\text{mol cm}^{-2} \text{yr}^{-1}$ )	Integration depth (D:L-modelled C-ox in sulphate zone (mbsf))	D:L-modelled C-ox in sulphate zone (bacterial dominance)* ( $\mu\text{mol cm}^{-2} \text{yr}^{-1}$ )	D:L-modelled C-ox in sulphate zone (archaeal dominance)† ( $\mu\text{mol cm}^{-2} \text{yr}^{-1}$ )	Integration depth (total D:L-modelled C-ox) (mbsf)	Total D:L-modelled C-ox (bacterial dominance)* ( $\mu\text{mol cm}^{-2} \text{yr}^{-1}$ )	Total D:L-modelled C-ox (archaeal dominance)† ( $\mu\text{mol cm}^{-2} \text{yr}^{-1}$ )
1227	427	1–34	1.8	5–35	3.9	4.0	5–85	6.6	6.7
1229	151	1–30	1.8 ‡	5–30	8.9	9.0	5–80	18.1	18.4
1230	5,086	1–10	5.0	5–10	1.2	1.2	5–200	21.3	21.4

\* It is assumed that all living cells belong to Bacteria

† It is assumed that 10% of all living cells belong to the Bacteria and 90% to the Archaea

‡ At site 1229 carbon oxidation (C-ox) by sulphate reduction was assumed to be similar to that at site 1227 because sulphate penetrated to approximately the same depth at the two sites.



surface world<sup>1</sup>. We estimate the minimum turnover time of the buried total organic carbon pool by dividing the total organic carbon concentration by the D:L-modelled carbon oxidation rates at all depths and sites. The total organic carbon turnover time is less than 43 million years, irrespective of whether bacteria or archaea dominate the microbial community. The total organic-carbon turnover times increase with approximately an order of magnitude from the upper and youngest sediment layers to the base of the cores. This suggests that buried organic carbon is indeed sufficient to fuel microbial activities over timescales of millions of years.

## METHODS SUMMARY

**Sediment handling and analysis.** Samples were obtained from sites 1227, 1229 and 1230 of ODP Leg 201 (ref. 23). Sub-samples were taken from the central part of the cores immediately after these were brought on deck. Samples were stored at  $-20^{\circ}\text{C}$  until analysis. Concentrations of muramic acid, D- and L-Asp, and individual amino acids were analysed by high-performance liquid chromatography (HPLC) following the method described in a previous study<sup>22</sup> with the exception that in cases in which one of the samples was spiked, muramic acid was analysed twice in each sample. Dipicolinic-acid concentrations were analysed by HPLC and calibrated by standard additions to samples<sup>24</sup>. Total organic carbon was determined in a Carlo Erba analyser after removal of carbonates by acidification with  $\text{H}_2\text{SO}_3$  (5–6% w/w). Total organic carbon was estimated from five-point standard curves and flour containing 44.39% carbon was used as a standard.

**Cell-specific concentrations of D- and L-amino acids.** Seven pure cultures, isolated from ODP Leg 201, were analysed for D- and L-Asp. Cultures were washed twice in saline before acid hydrolysis to remove amino acids from the growth medium. Cultures (2 ml) were added to 2 ml of  $12\text{ mol l}^{-1}$  HCl and hydrolysed at  $105^{\circ}\text{C}$  for 24 h under  $\text{N}_2$ . Concentrations of the respective D- and L-amino acids were analysed as described earlier. Cell numbers in the different cultures were determined by epifluorescence microscopy after DNA staining with 4'-6-diamidino-2-phenylindole.

**Standardization of data.** Samples for the present study were not obtained from exactly the same depths as AODC<sup>4</sup>. Hence, concentrations were read from profiles at the depth of AODC to make comparisons to AODC. No attempt was made to standardize profiles in the lower part of site 1229 because our samples did not cover the depths of elevated AODC. Moisture and density data were obtained from the IODP database (<http://iodp.tamu.edu/janusweb/physprops/maddat.cgi>).

Received 27 October 2011; accepted 27 January 2012.

Published online 18 March 2012.

1. D'Hondt, S. *et al.* Distributions of microbial activities in deep subseafloor sediments. *Science* **306**, 2216–2221 (2004).
2. Jørgensen, B. B. & D'Hondt, S. A starving majority deep beneath the seafloor. *Science* **314**, 932–934 (2006).
3. Schrenk, M. O., Huber, J. A. & Edwards, K. J. Microbial provinces in the subseafloor. *Annu. Rev. Mar. Sci.* **2**, 279–304 (2010).
4. Jørgensen, B. B., D'Hondt, S. & Miller, D. J. (eds) Leg 201 synthesis: controls on microbial communities in deeply buried sediments. *Proc. ODP Sci. Res.* 1–45 (2006).
5. Lipp, J. S., Morono, Y., Inagaki, F. & Hinrichs, K. U. Significant contribution of Archaea to extant biomass in marine subsurface sediments. *Nature* **454**, 991–994 (2008).
6. Schippers, A. *et al.* Prokaryotic cells of the deep sub-seafloor biosphere identified as living bacteria. *Nature* **433**, 861–864 (2005).
7. Biddle, J. F., Fitz-Gibbon, S., Schuster, S. C., Brenchley, J. E. & House, C. H. Metagenomic signatures of the Peru Margin subseafloor biosphere show a genetically distinct environment. *Proc. Natl Acad. Sci. USA* **105**, 10583–10588 (2008).
8. Fichtel, J., Koster, J., Rullkötter, J. & Sass, H. Spore dipicolinic acid contents used for estimating the number of endospores in sediments. *FEMS Microbiol. Ecol.* **61**, 522–532 (2007).

9. Madigan, M. T. & Martinko, J. M. *Brock Biology of Microorganisms* 11th edn (Prentice Hall, 2006).
10. Schichnes, D., Nemson, J. A. & Ruzin, S. E. Fluorescent staining method for bacterial endospores. *Microscope* **54**, 91–93 (2006).
11. Hsieh, L. K. & Vary, J. C. Germination and peptidoglycan solubilization in *Bacillus megaterium* spores. *J. Bacteriol.* **123**, 463–470 (1975).
12. Whitman, W. B., Coleman, D. C. & Wiebe, W. J. Prokaryotes: the unseen majority. *Proc. Natl Acad. Sci. USA* **95**, 6578–6583 (1998).
13. Parkes, R. J. *et al.* A quantitative study of microbial decomposition of biopolymers in Recent sediments from the Peru Margin. *Mar. Geol.* **113**, 55–66 (1993).
14. Hartgers, W. A. *et al.* Evidence for only minor contributions from bacteria sedimentary carbon. *Nature* **369**, 224–227 (1994).
15. Schleifer, K. H. & Kandler, O. Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriol. Rev.* **36**, 407–477 (1972).
16. Bada, J. L. Racemization of amino acids in nature. *Interdiscip. Sci. Rev.* **7**, 30–46 (1982).
17. Takano, Y., Sato, R., Kaneko, T., Kobayashi, K. & Marumo, K. Biological origin for amino acids in a deep subterranean hydrothermal vent, Toyoha mine, Hokkaido, Japan. *Org. Geochem.* **34**, 1491–1496 (2003).
18. Er, C., Nagy, B. & Riser, E. C. Analysis of muramic acid in holocene microbial environments by gas chromatography, electron impact, and fast atom bombardment mass spectrometry. *Geomicrobiol. J.* **5**, 57–78 (1987).
19. Heijnen, J. J. & van Dijken, J. P. In search of a thermodynamic description of biomass yields for the chemotrophic growth of microorganisms. *Biotechnol. Bioeng.* **39**, 833–858 (1992).
20. Leloup, J. *et al.* Sulfate-reducing bacteria in marine sediment (Aarhus Bay, Denmark): abundance and diversity related to geochemical zonation. *Environ. Microbiol.* **11**, 1278–1291 (2009).
21. Lomstein, B., Aa, Jørgensen, B. B., Schubert, C. J. & Niggemann, J. Amino acid biogeo- and stereochemistry in coastal Chilean sediments. *Geochim. Cosmochim. Acta* **70**, 2970–2989 (2006).
22. Lomstein, B., Aa, Jørgensen, B. B. & Langerhuus, A. T. Accumulation of prokaryotic remains during organic matter diagenesis in surface sediments off Peru. *Limnol. Oceanogr.* **54**, 1139–1151 (2009).
23. Shipboard Scientific Party. Leg 201 summary *Proc. ODP Init. Rep.* **201**, 1–81 (2003).
24. Lomstein, B. A. & Jørgensen, B. B. Pre column liquid chromatographic determination of dipicolinic acid from bacterial endospores. *Limnol. Oceanogr. Methods* (in the press).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** Bacterial cultures were provided by H. Cypionka. We thank members of the Leg 201 cruise for assistance in obtaining and processing samples. This research used samples and data provided by the Ocean Drilling Program ([http://www-odp.tamu.edu/publications/201\\_IR/201ir.htm](http://www-odp.tamu.edu/publications/201_IR/201ir.htm)). The ODP was sponsored by the US National Science Foundation and participating countries under the management of Joint Oceanographic Institutions. We thank R. O. Holm and L. Poulsen for technical assistance and guidance with high-performance liquid chromatographic analyses. We thank D. L. Valentine for comments and suggestions to improve the manuscript. Financial support was provided by the Max Planck Society, the Danish National Research Foundation, the Danish National Science Research Council, the Danish Agency for Science, Technology and Innovation, the Faculty of Science and Technology at the University of Aarhus, and the US National Science Foundation.

**Author Contributions** A.J.S. and B.Aa.L. developed ideas and performed the project planning. B.Aa.L. performed the analysis and data processing of total organic carbon, amino acid composition, dipicolinic acid and D- and L-amino acids. A.T.L. performed the muramic acid analysis and data processing. B.B.J. developed the mathematical formulation of the D:L model together with A.T.L. and B.Aa.L. Estimation of sulphate reduction rates from sulphate profiles was carried out by B.B.J. and S.D. The manuscript was written by B.Aa.L., A.T.L. and B.B.J. All authors participated in early stages of data interpretation and provided editorial comments on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to B.Aa.L. ([bente.lomstein@biology.au.dk](mailto:bente.lomstein@biology.au.dk)).

# Wild-type microglia arrest pathology in a mouse model of Rett syndrome

Noël C. Derecki<sup>1,2</sup>, James C. Cronk<sup>1,2,3</sup>, Zhenjie Lu<sup>1</sup>, Eric Xu<sup>1,4</sup>, Stephen B. G. Abbott<sup>5</sup>, Patrice G. Guyenet<sup>5</sup> & Jonathan Kipnis<sup>1,2,3</sup>

Rett syndrome is an X-linked autism spectrum disorder. The disease is characterized in most cases by mutation of the *MECP2* gene, which encodes a methyl-CpG-binding protein<sup>1–5</sup>. Although *MECP2* is expressed in many tissues, the disease is generally attributed to a primary neuronal dysfunction<sup>6</sup>. However, as shown recently, glia, specifically astrocytes, also contribute to Rett pathophysiology. Here we examine the role of another form of glia, microglia, in a murine model of Rett syndrome. Transplantation of wild-type bone marrow into irradiation-conditioned *Mecp2*-null hosts resulted in engraftment of brain parenchyma by bone-marrow-derived myeloid cells of microglial phenotype, and arrest of disease development. However, when cranial irradiation was blocked by lead shield, and microglial engraftment was prevented, disease was not arrested. Similarly, targeted expression of *MECP2* in myeloid cells, driven by *Lysm*<sup>cre</sup> on a *Mecp2*-null background, markedly attenuated disease symptoms. Thus, through multiple approaches, wild-type *Mecp2*-expressing microglia within the context of an *Mecp2*-null male mouse arrested numerous facets of disease pathology: lifespan was increased, breathing patterns were normalized, apnoeas were reduced, body weight was increased to near that of wild type, and locomotor activity was improved. *Mecp2*<sup>+/-</sup> females also showed significant improvements as a result of wild-type microglial engraftment. These benefits mediated by wild-type microglia, however, were diminished when phagocytic activity was inhibited pharmacologically by using annexin V to block phosphatidylserine residues on apoptotic targets, thus preventing recognition and engulfment by tissue-resident phagocytes. These results suggest the importance of microglial phagocytic activity in Rett syndrome. Our data implicate microglia as major players in the pathophysiology of this devastating disorder, and suggest that bone marrow transplantation might offer a feasible therapeutic approach for it.

The role of glia in Rett syndrome has recently been recognized<sup>7–9</sup>. *Mecp2*-null astrocytes were unable to support the normal dendritic ramification of wild-type neurons growing in culture<sup>7</sup>, and expression of wild-type *MECP2* protein in astrocytes of *Mecp2*-null hosts dramatically ameliorated disease pathology<sup>9</sup>. *Mecp2*-null microglia were reported to be toxic to neurons *in vitro* through production of high levels of glutamate<sup>10</sup>.

Microglia, the brain-resident macrophages, are of haematopoietic origin<sup>11</sup>. Still, the idea of repopulation of brain microglia from bone-marrow-derived cells in adult mice under normal physiological conditions is controversial<sup>12</sup>. However, when bone marrow transplantation is preceded by irradiation-mediated immune ablation, bone-marrow-derived cells with microglia-like morphology and phenotype (expressing low levels of CD45 and high levels of CD11b) are found in the brain<sup>13,14</sup>. Microglia have received increasing attention in the pathophysiology of several neurodegenerative and neuropsychiatric diseases<sup>14–18</sup>.

We first studied microglial function in the context of *Mecp2*<sup>-/-</sup> male mice. Males possess a single mutant X chromosome, and thus manifest

a severe phenotype that includes markedly retarded growth, apnoeas, tremor, impaired gait and locomotor function, and a postnatal life expectancy of approximately 8 weeks<sup>4,19</sup> (Fig. 1a, b and Supplementary Movie 1).

To address the role of hematopoietically derived cells in the pathophysiology of Rett, *Mecp2*<sup>-/-</sup> (*Mecp2*<sup>tm1.1Jae</sup> and *Mecp2*<sup>tm2Bird</sup>) mice were subjected to lethal split-dose irradiation at postnatal day (P)28 (the approximate age at which neurological signs appear<sup>4</sup>). Mice were then injected intravenously with syngeneic bone marrow from C57Bl/6J mice ubiquitously expressing green fluorescent protein (GFP). Control groups were injected with autologous (*Mecp2*<sup>-/-</sup>) bone marrow, or left naive. Surprisingly, the lifespan of *Mecp2*-null recipients of wild-type bone marrow (wild type → *Mecp2*<sup>-/-</sup>) was significantly extended compared with *Mecp2*<sup>-/-</sup> mice receiving autologous bone marrow (*Mecp2*<sup>-/-</sup> → *Mecp2*<sup>-/-</sup>) or to naive *Mecp2*<sup>-/-</sup> mice (Fig. 1b and Supplementary Movie 2). Our oldest living wild-type → *Mecp2*<sup>-/-</sup> mice were over 48 weeks of age (Supplementary Movie 3); most experimental mice were euthanized at the age of ~16 weeks for tissue analysis.

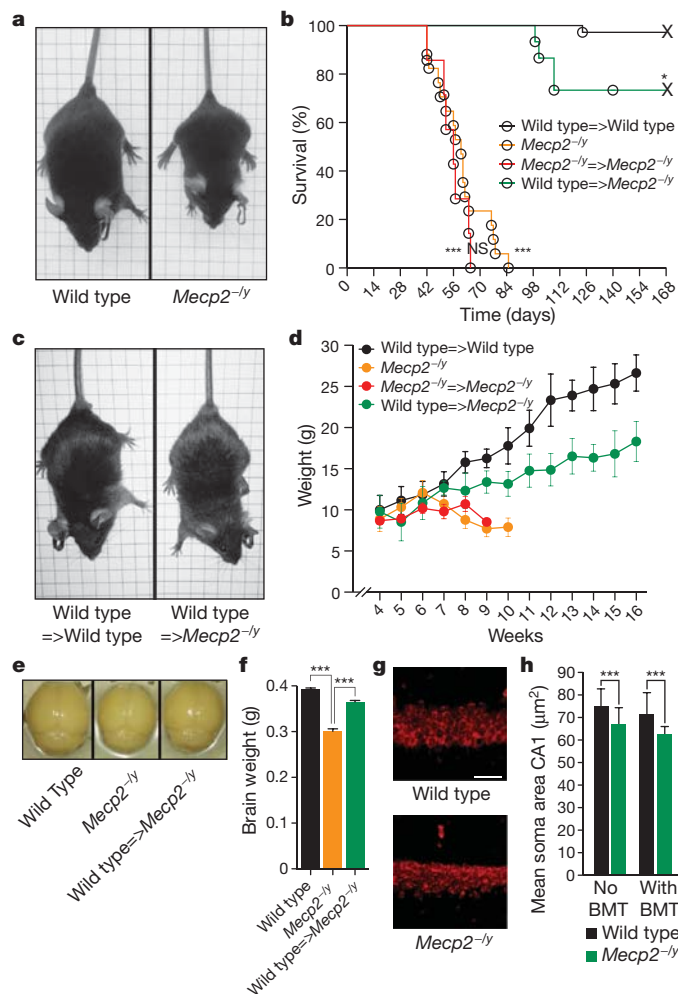
Although *Mecp2*<sup>-/-</sup> mice on the C57Bl/6J background are undersized<sup>4,19</sup>, within 4 weeks of transplantation, wild-type → *Mecp2*<sup>-/-</sup> (but not *Mecp2*<sup>-/-</sup> → *Mecp2*<sup>-/-</sup>) mice approached the size of wild-type littermates (Fig. 1c, d). Wild-type → *Mecp2*<sup>-/-</sup> mice also exhibited an increase in brain weight (Fig. 1e, f), which was probably caused by general growth of the mouse, because the reduced soma size characteristic of *Mecp2*-null neurons was not changed by bone marrow transplantation (Fig. 1g, h) and the brain to body weight ratio was normalized (Supplementary Fig. 1a). Spleens of *Mecp2*<sup>-/-</sup> mice were also smaller than normal and their size normalized after transplantation (Supplementary Fig. 1b–d).

Growth retardation is a characteristic feature of Rett pathology. Along these lines, treatment with insulin-like growth factor (IGF)-1 benefits survival and behavioural outcomes in *Mecp2*-null mice<sup>20</sup>. Indeed, peripheral macrophages from wild-type mice expressed significantly higher levels of IGF-1 *in vitro* in response to immunological stimuli than macrophages from *Mecp2*-null (*Mecp2*<sup>tm1.1Jae/y</sup>) mice (Supplementary Fig. 2); this difference may contribute to the increased body growth seen in *Mecp2*-null mice after bone marrow transplantation.

The general appearance of wild-type → *Mecp2*<sup>-/-</sup> mice was improved compared with that of naive *Mecp2*<sup>-/-</sup> or *Mecp2*<sup>-/-</sup> → *Mecp2*<sup>-/-</sup> mice. The severe involuntary tremors normally seen in mutant mice were absent following wild-type bone marrow transplantation (Fig. 2a), and gait was improved. Interestingly, no detectable benefit on hindlimb clasping phenotype was observed.

Breathing irregularities and apnoeas are cardinal signs of Rett syndrome. We used whole-body plethysmography to compare the breathing patterns of *Mecp2*<sup>-/-</sup> mice with or without bone marrow transplantation to those of control mice (Fig. 2b). As expected, *Mecp2*<sup>-/-</sup> mice developed apnoeas progressively with age (data not shown). However, wild type → *Mecp2*<sup>-/-</sup> exhibited significantly reduced apnoeas and fewer breathing irregularities than either naive *Mecp2*<sup>-/-</sup> or

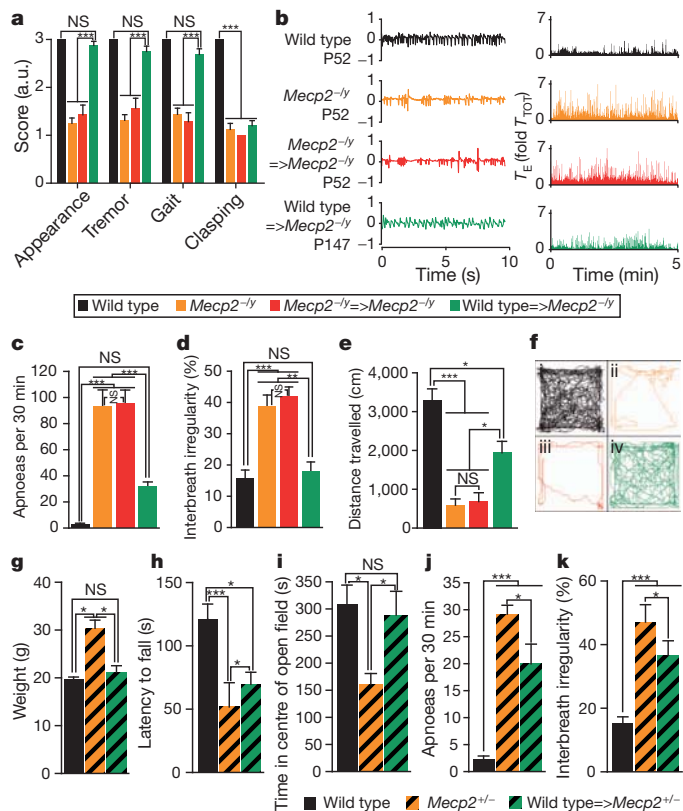
<sup>1</sup>Department of Neuroscience, School of Medicine, University of Virginia, Charlottesville, Virginia 22908, USA. <sup>2</sup>Graduate Program in Neuroscience, School of Medicine, University of Virginia, Charlottesville, Virginia 22908, USA. <sup>3</sup>Medical Scientist Training Program, School of Medicine, University of Virginia, Charlottesville, Virginia 22908, USA. <sup>4</sup>Undergraduate School of Arts and Sciences, University of Virginia, Charlottesville, Virginia 22908, USA. <sup>5</sup>Department of Pharmacology, School of Medicine, University of Virginia, Charlottesville, Virginia 22908, USA.



**Figure 1 | Wild-type bone marrow transplantation (BMT) at P28 arrests disease progression in *Mecp2*<sup>-/-</sup> mice.** **a**, Representative images of wild-type and *Mecp2*<sup>-/-</sup> littermates at P56. **b**, Lifespan of *Mecp2*<sup>-/-</sup> mice receiving wild-type bone marrow at P28 (wild type → *Mecp2*<sup>-/-</sup>; *n* = 15) compared with naive *Mecp2*<sup>-/-</sup> (*n* = 17), *Mecp2*<sup>-/-</sup> receiving *Mecp2*<sup>-/-</sup> bone marrow (*Mecp2*<sup>-/-</sup> → *Mecp2*<sup>-/-</sup>; *n* = 9) and wild-type mice receiving wild-type bone marrow (wild type → wild type; *n* = 29) (\*\**P* < 0.0001, log rank (Mantel-Cox)). **c**, Representative images of wild-type → wild-type compared with wild-type → *Mecp2*<sup>-/-</sup> mice are shown at P56 (4 weeks after bone marrow transplantation). **d**, Weights (mean ± s.e.m.) of wild-type → wild-type, *Mecp2*<sup>-/-</sup>, *Mecp2*<sup>-/-</sup> → *Mecp2*<sup>-/-</sup> and wild-type → *Mecp2*<sup>-/-</sup> mice (*n* = 15, 15, 7, 15 mice per group) are shown over time. **e**, Representative images of brains isolated from P56 wild-type → wild-type and wild-type → *Mecp2*<sup>-/-</sup> mice transplanted at P28 and naive *Mecp2*<sup>-/-</sup> mice. **f**, Brain weight (mean ± s.e.m.) for each group (\*\**P* < 0.001; one-way analysis of variance (ANOVA); *n* = 4 each group). **g**, Representative images of Nissl staining in hippocampal slices (CA1 area) of wild-type and *Mecp2*<sup>-/-</sup> mice (scale bar, 40 μm). **h**, Soma area (mean ± s.d.) of CA1 hippocampal cells. For each group of mice, a set of cells was created by randomly selecting 100 cells per mouse, three mice per group (\*\**P* < 0.001; two-way ANOVA with Bonferroni *post hoc* test).

*Mecp2*<sup>-/-</sup> → *Mecp2*<sup>-/-</sup> mice (Fig. 2c, d). Wild-type → *Mecp2*<sup>-/-</sup> mice also displayed significantly increased mobility in the open field compared with naive *Mecp2*<sup>-/-</sup> or *Mecp2*<sup>-/-</sup> → *Mecp2*<sup>-/-</sup> mice (Fig. 2e, f).

We also performed bone marrow transplantation in heterozygous female mice at 2 months of age, and animals were examined at 9 months. The disease in *Mecp2*<sup>+/-</sup> mice develops slowly, with behavioural abnormalities becoming clear at 4–6 months of age. Weights of treated *Mecp2*<sup>+/-</sup> mice were comparable to wild-type controls (Fig. 2g). Moreover, there was significant improvement in motor function, as examined on rotarod (Fig. 2h), and time spent in the centre of the open field arena was significantly increased (Fig. 2i).



**Figure 2 | Bone marrow transplantation effects on general appearance, breathing and locomotion of *Mecp2*<sup>-/-</sup> and *Mecp2*<sup>+/-</sup> mice.** **a**, Neurological scores at P56 for wild-type → wild-type, *Mecp2*<sup>-/-</sup> naive, *Mecp2*<sup>-/-</sup> → *Mecp2*<sup>-/-</sup> and wild-type → *Mecp2*<sup>-/-</sup> mice. Behaviours (mean ± s.e.m.) are scored as indicated in Methods (\*\**P* < 0.001; one-way ANOVA; *n* = 16, 16, 7, 16); NS, not significant; a.u., arbitrary units. **b**, Left, representative plethysmograph recordings of animals from each group; right, expiratory time (*T<sub>E</sub>*) for representative wild-type, *Mecp2*<sup>-/-</sup>, *Mecp2*<sup>-/-</sup> → *Mecp2*<sup>-/-</sup> and wild-type → *Mecp2*<sup>-/-</sup> mice (transplantation at P28 and examination at indicated ages for all groups) as measured over 5 min; *T<sub>E</sub>* is normalized to mean breath duration for each mouse. **c**, Apnoeas (mean ± s.e.m.) per 30 min as measured in all four groups (\*\**P* < 0.001; one-way ANOVA with Bonferroni *post hoc* test; *n* = 5 mice per group; for the entire figure, all mice were age P56 except for wild-type → *Mecp2*<sup>-/-</sup> at 12 weeks of age; that is, 8 weeks after bone marrow transplantation). **d**, Interbreath irregularity (mean percentage ± s.e.m.) calculated as absolute ( $(T_{\text{total},n} - T_{\text{total},n+1})/T_{\text{total},n+1}$ ) for all four groups (\*\**P* < 0.01; \*\*\**P* < 0.001; one-way ANOVA with Bonferroni *post hoc* test; *n* = 5 mice per group). **e**, Distance travelled (mean ± s.e.m.) in an open field (\**P* < 0.05; \*\*\**P* < 0.001; one-way ANOVA, *n* = 5 mice per group). **f**, Representative traces of the path travelled by mice in an open field during 20 min test time. **g–k**, *Mecp2*<sup>+/-</sup> mice were transplanted with wild-type bone marrow at P56 and were examined for disease symptoms at 9 months of age. **g**, Weight (mean ± s.e.m.); **h**, latency to fall (mean ± s.e.m.) in the rotarod task; **i**, time (mean ± s.e.m.) spent in the centre of the open field; **j**, apnoeas (mean ± s.e.m.) measured by whole-body plethysmography in 30 min; **k**, interbreath irregularity (mean ± s.e.m.): all were improved in the treated mice compared with non-treated controls (\**P* < 0.05; \*\*\**P* < 0.001; one-way ANOVA, *n* = 6 mice per group; *post hoc* Bonferroni test was used for individual comparisons).

Apnoeas in bone marrow transplanted mice were reduced (Fig. 2j) and their overall breathing patterns were improved compared with their non-treated controls (Fig. 2k).

The peripheral immune system of *Mecp2*<sup>-/-</sup> hosts was repopulated by donor bone marrow (Supplementary Fig. 3a). Additionally, it has been previously shown that bone marrow transplantation after whole-body irradiation results in engraftment of microglia-like myeloid cells into the brain parenchyma<sup>13</sup>. Indeed, GFP<sup>+</sup> cells in the parenchyma of bone marrow transplanted mice expressed CD11b (Fig. 3a) but not GFAP or NeuN (data not shown). Twelve weeks after bone marrow transplantation, mice were perfused and their brains dissected into



sub-areas, prepared in single-cell suspensions and analysed using flow cytometry to determine percentages of bone-marrow-derived (GFP<sup>+</sup>) cells in the haematopoietic (CD45<sup>+</sup>) cell fractions in the brain (Supplementary Fig. 3b, c).

Interestingly, in mice in which bone marrow transplantation was performed later (P40 or P45), only slight improvements in disease pathology were observed (Supplementary Fig. 4a). No microglial engraftment was evident, although substantial numbers of GFP<sup>+</sup> cells were found in the meningeal spaces (Supplementary Fig. 4b). These results may suggest that when disease progression is faster than microglial engraftment, full rescue cannot be achieved. The moderate results observed, however, may have been due to a yet-unknown mechanism, perhaps through soluble factors produced by meningeal immune cells, or peripherally-expressed IGF-1 (Supplementary Fig. 2). When bone marrow transplantation was performed at P2 without irradiation, minimal peripheral chimaerism was achieved without detectable microglial engraftment and no lifespan extension was observed (Supplementary Fig. 4a, c).

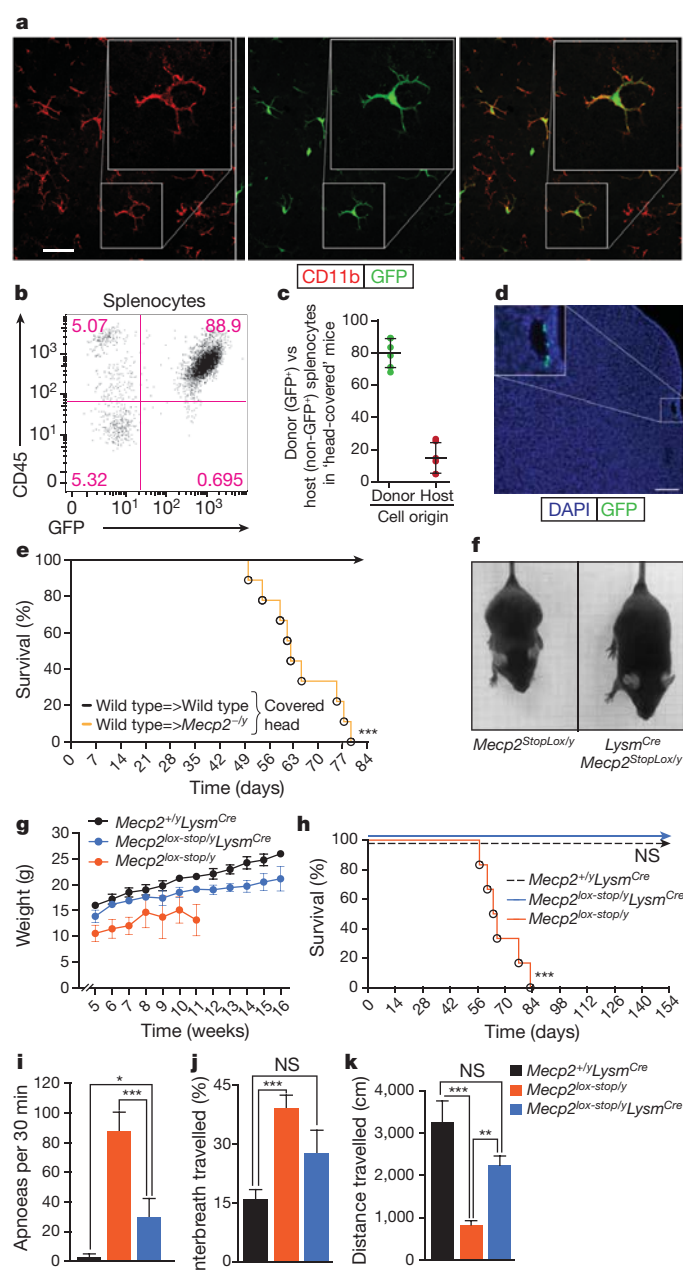
Newly engrafted microglia expressed detectable levels of wild-type *Mecp2* (data not shown) but nearby cells did not show any *Mecp2*

labelling, arguing against the possibility of protein or messenger RNA transfer from engrafted microglia into nearby cells as an underlying mechanism for the beneficial effect of bone marrow transplantation.

To substantiate the specific role of microglia in bone marrow transplantation-mediated disease arrest, we repeated transplantation experiments, again at P28, but with the addition of lead shielding to block cranial irradiation, which results in repopulation of peripheral immunity (Fig. 3b, c) but no parenchymal engraftment (Fig. 3d), supporting previously published work<sup>13,18</sup>. Disease was not arrested in 'head-covered' mice (Fig. 3e), suggesting that peripheral immune reconstitution without microglial engraftment is insufficient to arrest pathology in *Mecp2*<sup>-/-</sup> mice.

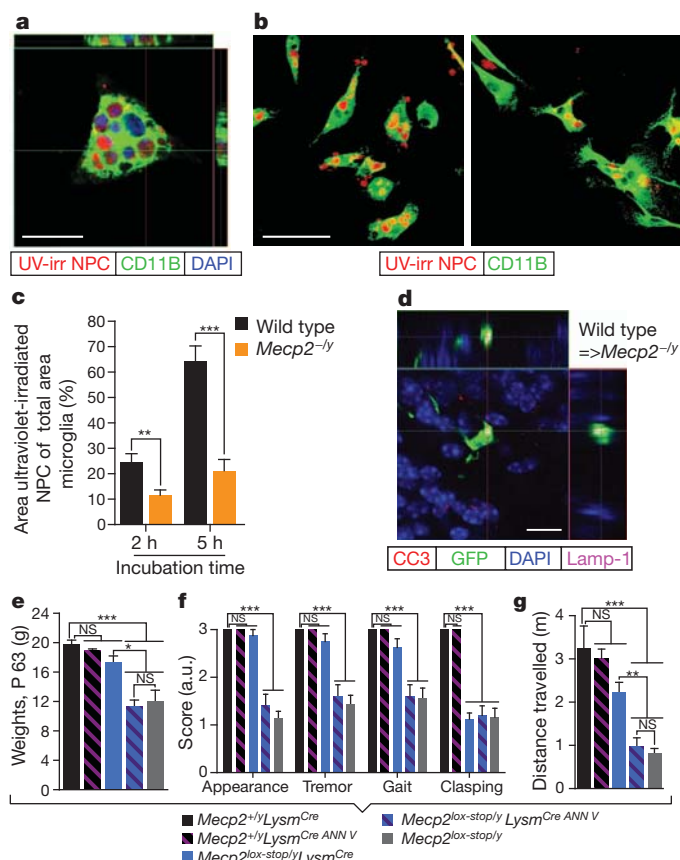
To substantiate further the role of myeloid cells in arrest of Rett pathology, we used a genetic approach. We employed the widely used *Lysm*<sup>Cre</sup> mouse—which results in a high degree of recombination in myeloid cells, granulocytes and in significant numbers of microglia<sup>21–23</sup>—in cross with *Mecp2*<sup>lox-stop</sup> mice. Male progeny, *Mecp2*<sup>lox-stop/y</sup>*Lysm*<sup>Cre</sup>, express wild-type *Mecp2* in myeloid cells on an otherwise *Mecp2*-null background. These animals exhibited improvements in overall appearance and growth (Fig. 3f, g and Supplementary Movie 4) and their lifespans were significantly increased (Fig. 3h). The oldest *Mecp2*<sup>lox-stop/y</sup>*Lysm*<sup>Cre</sup> animals were 31 weeks of age, with survival of 100%; *n* = 6 mice per group. Apnoeas and interbreath irregularity of these mice were also significantly reduced compared with control mice (Fig. 3i, j), and their open field activity was not significantly different from wild-type counterparts (Fig. 3k). These results cannot be interpreted by cre leakiness, because no cre-mediated recombination was evident in either astrocytes or neurons in *Lysm*<sup>Cre</sup> crossed to a reporter strain (data not shown), in line with previous publications<sup>22,23</sup>.

Microglia from *Mecp2*-null mice were deficient in their response to immunological stimuli (Supplementary Fig. 5) and in phagocytic capacity, as examined by feeding cultured microglia with pre-labelled ultraviolet-irradiated neural progenitor cells, used as apoptotic targets<sup>24</sup> (Fig. 4a–c). Thus, it is possible that apoptotic debris would accumulate over time in the *Mecp2*-null brain, contributing to neuronal malfunction and accelerating disease progression. Along these lines, supplementation of wild-type microglia could reduce debris levels and allow improved neuronal function. Indeed, in mice transplanted with GFP<sup>+</sup>



**Figure 3 | Brain parenchymal engraftment of cells after bone marrow transplantation is required to arrest Rett syndrome.**

**a**, Representative confocal images of brain parenchyma from cerebellum of wild-type → *Mecp2*<sup>-/-</sup> mice 8 weeks after transplantation (at P28), immunolabelled for CD11b and GFP (scale bar, 20 μm). **b–e**, *Mecp2*<sup>-/-</sup> mice underwent bone marrow transplantation at P28 with their heads lead-protected. Mice were examined at their end-point, about 7 weeks after bone marrow transplantation. **b**, Representative dot plot of splenocytes obtained from bone-marrow-transplanted mouse with lead-protected head. **c**, Distribution of 'peripheral chimaerism' in mice with lead-protected heads after bone marrow transplantation. **d**, Representative micrograph from mice with lead-protected heads after bone marrow transplantation, immunolabelled for GFP. Coronal cortical slice is presented showing sporadic cells found in meningeal spaces, but not in the parenchyma. **e**, Lifespan of *Mecp2*<sup>-/-</sup> mice with wild-type bone marrow transplantation with lead-covered heads compared with wild-type → wild-type controls with lead-covered heads (\*\*\*) *P* < 0.0001, log rank (Mantel-Cox); *n* = 9 mice per group). **f–k**, Genetic approach for expressing MECP2 protein in myeloid cells. *Mecp2*<sup>lox-stop</sup> mice were bred to *Lysm*<sup>Cre</sup> mice and their progeny (*Mecp2*<sup>lox-stop/y</sup>*Lysm*<sup>Cre</sup> mice) were analysed for disease progression. **f**, Representative image of mice at P56. **g**, Weights (mean ± s.e.m.) of mice as they progress with age. **h**, Lifespan for indicated groups (\*\*\*) *P* < 0.0001, log rank (Mantel-Cox); *n* = 6 mice per group). **i**, Apnoeas (mean ± s.e.m.) measured by whole-body plethysmography in 30 min for the three groups at 9 weeks. **j**, Interbreath irregularity (mean % ± s.e.m.) measured at 9 weeks. **k**, Distance travelled (mean ± s.e.m.) in an open field at 9 weeks (\*\*\*) *P* < 0.01; \*\*\*) *P* < 0.001; one-way ANOVA, *n* = 5 mice per group; Bonferroni *post hoc* test was used for individual comparisons).



**Figure 4 | Microglial phagocytic activity is necessary for their beneficial effect in *Mecp2<sup>-ly</sup>* mouse brains.** **a**, Representative micrograph of phagocytosing microglia in orthogonal projections of confocal z-stacks. Scale bar, 25  $\mu$ m. **b**, Wild-type (left) or *Mecp2<sup>-ly</sup>* (right) microglia incubated for 2 or 5 h with fluorescently labelled ultraviolet-irradiated neural progenitor cells and stained with anti-CD11b. Scale bar, 50  $\mu$ m. **c**, Bar graphs comparing surface area of ultraviolet-irradiated neural progenitor cells (NPC) to total surface area (mean  $\pm$  s.e.m.) of wild-type or *Mecp2<sup>-ly</sup>* microglia (\*\* $P < 0.01$ ; \*\*\* $P < 0.001$ ; one-way ANOVA; representative experiment shown out of three independently performed). **d**, Representative micrograph of phagocytosing microglia *in situ* containing cleaved caspase-3 debris co-localized with lysosomal marker, Lamp-1. Scale bar, 50  $\mu$ m. **e–g**, *Mecp2<sup>lox-stop/y</sup>Lysm<sup>Cre</sup>* mice and the appropriate controls were treated with annexin V to inhibit phagocytic activity pharmacologically. **e**, Weights (mean  $\pm$  s.e.m.) of *Mecp2<sup>+/y</sup>Lysm<sup>Cre</sup>*, *Mecp2<sup>lox-stop/y</sup>Lysm<sup>Cre</sup>*, *Mecp2<sup>lox-stop/y</sup>Lysm<sup>Cre</sup> ANN V* and *Mecp2<sup>lox-stop/y</sup>Lysm<sup>Cre</sup>* mice treated with annexin V are shown at the end point for *Mecp2<sup>lox-stop/y</sup>Lysm<sup>Cre</sup>* and *Mecp2<sup>lox-stop/y</sup>Lysm<sup>Cre</sup>* treated with annexin V groups ( $\sim$ P63). **f**, Neurological scores (mean  $\pm$  s.e.m.) at P56 are presented (\*\* $P < 0.001$ ; one-way ANOVA;  $n$  = at least seven mice per group; Bonferroni *post hoc* test was used for individual comparisons). **g**, Distance travelled (mean  $\pm$  s.e.m.) in an open field by mice from all the above groups (\*\* $P < 0.001$ ; one-way ANOVA,  $n$  = 5 mice per group; Bonferroni *post hoc* test was used for individual comparisons).

bone marrow, only GFP<sup>+</sup> parenchymal cells were consistently found containing cleaved caspase-3-positive debris within lysosomes (Fig. 4d).

It has been previously shown that annexin V (a protein that binds phosphatidylserine on apoptotic cells and inhibits engulfment) injected intravenously can reach the central nervous system<sup>25</sup>. Moreover, we have recently shown that intravenous injection of annexin V results in substantial blockade of phagocytic activity in the brain<sup>24</sup>. Indeed, treatment of wild-type mice with annexin V resulted in significant accumulation of terminal deoxynucleotidyl transferase (TdT)-mediated dUTP nick end labelling (TUNEL)<sup>+</sup> fragments (Supplementary Fig. 6).

We attempted to inhibit brain phagocyte activity pharmacologically in *Mecp2<sup>lox-stop/y</sup>Lysm<sup>Cre</sup>* mice and compare disease progression with

controls. Long-term treatment of *Mecp2<sup>lox-stop/y</sup>Lysm<sup>Cre</sup>* mice with annexin V abolished the amelioration of the disease normally seen in these mice (Fig. 4e–g). Wild-type mice treated with annexin V were not significantly affected. This is probably because, unlike in *Mecp2<sup>lox-stop/y</sup>Lysm<sup>Cre</sup>* mice, neurons and astrocytes in wild-type mice are fully functional, expressing wild-type MECP2. It is conceivable, however, that a longer treatment of wild-type mice with annexin V might result in neurological pathology. Overall, these results suggest active engagement of wild-type microglia in clearance of apoptotic cells or cell remnants within the context of otherwise *Mecp2*-null brain—a task that probably cannot be sufficiently performed by *Mecp2*-null microglia.

Neuropathologists have observed gliosis and cell loss in the cerebellum of deceased patients who had Rett syndrome<sup>26</sup>, but this work has not received much attention, presumably because the disease is generally considered non-neurodegenerative. Our results do not claim that neurodegeneration underlies the pathology of the disease. Rather, they suggest that *Mecp2*-null microglia, deficient in phagocytic function, may be unable to keep pace in clearing debris left behind from the normal process of neural cell death or membrane shedding. This, in turn, would lead to a crowded and sub-optimal central nervous system milieu within which neurons, already challenged by loss of *Mecp2*, might be further impaired in function. The inability of *Mecp2*-null microglia to clear debris as effectively as wild-type microglia has the potential to contribute to the underlying neuropathology and/or the time course of appearance of symptoms in *Mecp2*-null mice<sup>4,27</sup>.

Future studies should be aimed at understanding the connections between glial phagocytic activity and neuronal function, and possible interactions between microglia and astrocytes in the pathology of Rett syndrome. Phagocytic activity per se is almost certainly just one aspect of glial involvement in the pathophysiology of the disease. It is conceivable that glia, including astrocytes—which are also capable phagocytes<sup>28</sup>—release soluble factors in connection with their own phagocytic activity, in turn benefiting neuronal function. Therefore, removal of debris itself may not be as primarily relevant to disease progression as a secondary response of glia to the phagocytic process. Accordingly, inhibition of phagocytosis might result in exacerbation of pathology of the disease through these as yet unknown processes, even in the absence of deposits of easily observable cellular debris.

Our present findings support previous publications describing the potential for clinical treatment of Rett pathology<sup>6,29</sup>, while also suggesting the possibility of achieving this goal through augmentation or repopulation of brain phagocytes, or improvement of their phagocytic activity. These results open the possibility for a new approach in the amelioration of the pathology of Rett syndrome.

## METHODS SUMMARY

**Animals.** Male and female C57Bl/6-Tg(UBC-GFP)30Scha/J, C57Bl/6J, B6.129P2(C)Mecp2<sup>tm1.1Bird</sup>/J, B6.129P2-Lyz2<sup>tm1(cres)</sup>/J (*Lysm<sup>Cre</sup>*), B6.129P2-Mecp2<sup>tm2Bird</sup>/J, *Mecp2<sup>lox-stop/y</sup>* and C57Bl/6J mice were purchased from Jackson Laboratories; B6.Cg-Mecp2<sup>tm1.1Jae/Mmcd</sup> mice were a gift from A. Pieper and maintained in our laboratory on C57Bl/6J background. All procedures complied with regulations of the Institutional Animal Care and Use Committee at The University of Virginia.

**Irradiation and bone marrow transfer.** Four-week-old mice were subjected to lethal split-dose  $\gamma$ -irradiation (300 rad followed 48 h later by 950 rad). Four hours after the second irradiation, mice were injected with  $5 \times 10^6$  bone marrow cells. After irradiation, mice were kept on drinking water fortified with sulphamethoxazole for 2 weeks to limit infection by opportunistic pathogens.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 3 May 2011; accepted 27 January 2012.

Published online 18 March 2012.

1. Van den Veyver, I. B. & Zoghbi, H. Y. Mutations in the gene encoding methyl-CpG-binding protein 2 cause Rett syndrome. *Brain Dev.* **23** (suppl. 1), S147–S151 (2001).

2. Van den Veyver, I. B. & Zoghbi, H. Y. Genetic basis of Rett syndrome. *Ment. Retard. Dev. Disabil. Res. Rev.* **8**, 82–86 (2002).
3. Amir, R. E. *et al.* Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nature Genet.* **23**, 185–188 (1999).
4. Guy, J., Hendrich, B., Holmes, M., Martin, J. E. & Bird, A. A mouse *Mecp2*-null mutation causes neurological symptoms that mimic Rett syndrome. *Nature Genet.* **27**, 322–326 (2001).
5. Nan, X. & Bird, A. The biological functions of the methyl-CpG-binding protein MeCP2 and its implication in Rett syndrome. *Brain Dev.* **23** (suppl. 1), S32–S37 (2001).
6. Luikenhuis, S., Giacometti, E., Beard, C. F. & Jaenisch, R. Expression of MeCP2 in postmitotic neurons rescues Rett syndrome in mice. *Proc. Natl Acad. Sci. USA* **101**, 6033–6038 (2004).
7. Ballas, N., Lioy, D. T., Grunseich, C. & Mandel, G. Non-cell autonomous influence of MeCP2-deficient glia on neuronal dendritic morphology. *Nature Neurosci.* **12**, 311–317 (2009).
8. Maezawa, I., Swanberg, S., Harvey, D., LaSalle, J. M. & Jin, L. W. Rett syndrome astrocytes are abnormal and spread MeCP2 deficiency through gap junctions. *J. Neurosci.* **29**, 5051–5061 (2009).
9. Lioy, D. T. *et al.* A role for glia in the progression of Rett's syndrome. *Nature* **475**, 497–500 (2011).
10. Maezawa, I. & Jin, L. W. Rett syndrome microglia damage dendrites and synapses by the elevated release of glutamate. *J. Neurosci.* **30**, 5346–5356 (2010).
11. Ginhoux, F. *et al.* Fate mapping analysis reveals that adult microglia derive from primitive macrophages. *Science* **330**, 841–845 (2010).
12. Ajami, B., Bennett, J. L., Krieger, C., Tetzlaff, W. & Rossi, F. M. Local self-renewal can sustain CNS microglia maintenance and function throughout adult life. *Nature Neurosci.* **10**, 1538–1543 (2007).
13. Mildner, A. *et al.* Microglia in the adult brain arise from Ly-6ChiCCR2<sup>+</sup> monocytes only under defined host conditions. *Nature Neurosci.* **10**, 1544–1553 (2007).
14. Boissonneault, V. *et al.* Powerful beneficial effects of macrophage colony-stimulating factor on  $\beta$ -amyloid deposition and cognitive impairment in Alzheimer's disease. *Brain* **132**, 1078–1092 (2009).
15. Chen, S. K. *et al.* Hematopoietic origin of pathological grooming in *Hoxb8* mutant mice. *Cell* **141**, 775–785 (2010).
16. Hoogerbrugge, P. M. *et al.* Donor-derived cells in the central nervous system of twitcher mice after bone marrow transplantation. *Science* **239**, 1035–1038 (1988).
17. Simard, A. R., Soulet, D., Gowing, G., Julien, J. P. & Rivest, S. Bone marrow-derived microglia play a critical role in restricting senile plaque formation in Alzheimer's disease. *Neuron* **49**, 489–502 (2006).
18. Shechter, R. *et al.* Infiltrating blood-derived macrophages are vital cells playing an anti-inflammatory role in recovery from spinal cord injury in mice. *PLoS Med.* **6**, e1000113 (2009).
19. Chen, R. Z., Akbarian, S., Tudor, M. & Jaenisch, R. Deficiency of methyl-CpG binding protein-2 in CNS neurons results in a Rett-like phenotype in mice. *Nature Genet.* **27**, 327–331 (2001).
20. Tropea, D. *et al.* Partial reversal of Rett syndrome-like symptoms in MeCP2 mutant mice. *Proc. Natl Acad. Sci. USA* **106**, 2029–2034 (2009).
21. Willemsen, H. L. *et al.* Microglial/macrophage GRK2 determines duration of peripheral IL-1 $\beta$ -induced hyperalgesia: contribution of spinal cord CX3CR1, p38 and IL-1 signaling. *Pain* **150**, 550–560 (2010).
22. Nijboer, C. H. *et al.* Cell-specific roles of GRK2 in onset and severity of hypoxic-ischemic brain damage in neonatal mice. *Brain Behav. Immun.* **24**, 420–426 (2010).
23. Cho, I. H. *et al.* Role of microglial IKK $\beta$  in kainic acid-induced hippocampal neuronal cell death. *Brain* **131**, 3019–3033 (2008).
24. Lu, Z. *et al.* Phagocytic activity of neuronal progenitors regulates adult neurogenesis. *Nature Cell Biol.* **13**, 1076–1083 (2011).
25. Zhang, X. *et al.* A minimally invasive, translational biomarker of ketamine-induced neuronal death in rats: microPET imaging using 18F-annexin V. *Toxicol. Sci.* **111**, 355–361 (2009).
26. Oldfors, A. *et al.* Rett syndrome: cerebellar pathology. *Pediatr. Neurol.* **6**, 310–314 (1990).
27. McGraw, C. M., Samaco, R. C. & Zoghbi, H. Y. Adult neural function requires MeCP2. *Science* **333**, 186 (2011).
28. Park, D. *et al.* BAI1 is an engulfment receptor for apoptotic cells upstream of the ELMO/Dock180/Rac module. *Nature* **450**, 430–434 (2007).
29. Guy, J., Gan, J., Selfridge, J., Cobb, S. & Bird, A. Reversal of neurological defects in a mouse model of Rett syndrome. *Science* **315**, 1143–1147 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank S. Smith for editing the manuscript. We thank the members of the Kipnis laboratory as well as the members of the University of Virginia Neuroscience Department for their comments during multiple discussions of this work. We also thank S. Feldman for injection of neonatal mice, I. Smirnov for tail vein injections, and B. Tomlin and J. Jones for their animal care. N.C.D. is a recipient of a Hartwell Foundation post-doctoral fellowship. This work was primarily supported by a grant from the Rett Syndrome Research Trust (to J.K.) and in part by HD056293 and AG034113 (to J.K.).

**Author Contributions** N.C.D. performed most of the experiments, analysed the data and prepared it for presentation, and contributed to experimental design and manuscript writing. J.C.C. performed the experiments with phagocytic activity of microglia *in vivo* and assisted with additional immunohistochemistry experiments along with data analysis and presentation, and contributed to experimental design and manuscript editing. Z.L. assisted with *in vitro* phagocytic activity experiments. E.X. assisted with animal behaviour scoring. S.B.G.A. assisted with plethysmography experiments and data analysis. P.G.G. assisted with plethysmography experimental design, data analysis and presentation of plethysmography-related data, and contributed to manuscript editing. J.K. designed the study, assisted with data analysis and presentation, and wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to J.K. ([kipnis@virginia.edu](mailto:kipnis@virginia.edu)).



## METHODS

**Animals.** Male and female C57Bl/6-Tg(UBC-GFP)30Scha/J, C57Bl/6J, B6.129P2(C)Mecp2<sup>tm1.1Bird</sup>/J, B6.129P2-Lyz2<sup>tm1.1cre</sup>/Jf/J (*Lysm*<sup>cre</sup>), B6.129P2-Mecp2<sup>tm2Bird</sup>/J, *Mecp2*<sup>lox-stop/y</sup> and C57Bl/6J mice were purchased from Jackson Laboratories; B6.Cg-Mecp2<sup>tm1.1Jae/Mmcd</sup> mice were a gift from A. Pieper (Southwestern Medical School, Dallas, Texas, USA) and were maintained in our laboratory on C57Bl/6J background. All animals were housed in temperature and humidity controlled rooms, maintained on a 12 h/12 h light/dark cycle (lights on at 7:00) and age-matched in each experiment. All strains were kept in identical housing conditions. All procedures complied with regulations of the Institutional Animal Care and Use Committee at The University of Virginia.

**Neurological assays.** Mice were weighed on a laboratory scale and weights recorded to the nearest tenth of a gram; all other assays were scored on a scale from 3 to 1, with '3' being wild type (normal), and '1' being severe, as follows: gait '1', wide-spread hind limbs, severe 'waddling gait,' hopping or (unintentional) reverse locomotion; claspings '1', clenching of both rear hindlimbs tightly across ventral body aspect; tremor '1', visible involuntary shaking, particularly during or after handling; appearance '1', greasy and/or unkempt fur, clear signs of severe self-neglect in terms of grooming; kyphosis, hunched posture; eyes sunken, partly or fully closed, watery or crusted over. Scores of '2' were assigned to any phenotypes in between wild type and severe.

**Bone marrow isolation.** Mice were killed using CO<sub>2</sub> saturated with 70% alcohol. Skin was removed from the lower part of the body. Tissue was removed from hindlimbs with scissors and dissected away from body. Remaining tissue was cleaned from the tibial and femoral bones and bones were separated at the knee joint. Bone ends were cut off. Cells were forced out of bones with a stream of pH 7.4 0.1 M PBS containing 10% fetal calf serum using a 10 ml syringe with 25 gauge needle. Cells were centrifuged and re-suspended at the concentration of  $2 \times 10^7$  cells per millilitre in PBS (250  $\mu$ l cell suspension was injected intravenously in each animal through the tail vein).

**Irradiation and bone marrow transfer.** Four-week-old wild-type C57Bl/6J, B6.129P2(C)Mecp2<sup>tm1.1Bird</sup>–/y, or B6.Cg-Mecp2<sup>tm1.1Jae/Mmcd</sup>–/y mice were subjected to lethal split-dose  $\gamma$ -irradiation (300 rad followed 48 h later by 950 rad). Four hours after the second irradiation, mice were injected with  $5 \times 10^6$  bone-marrow cells freshly isolated from C57Bl/6-Tg(UBC-GFP)30Scha/J, CBySmn.CB17-Prkdc<sup>scid</sup>/J or B6.Cg-Mecp2<sup>tm1.1Jae/Mmcd</sup>–/y mice. After irradiation, mice were kept on drinking water fortified with sulphamethoxazole for 2 weeks to limit infection by opportunistic pathogens.

**Fluorescence-activated cell sorting (FACS) of brain isolates.** Mice were perfused with 0.1 M PBS for 5 min. Heads were removed and skulls were quickly stripped of all flesh. Mandibles were next removed, as was all skull material rostral to maxillae. Surgical scissors (Fine Science Tools) were used to remove tops of skulls, cutting clockwise, beginning and ending inferior to the right post-tympanic hook. Brains were immediately placed in ice-cold FACS buffer (pH 7.4 0.1 M PBS, 1 mM EDTA, 1% BSA). Meninges (dura mater, arachnoid mater and pia mater) were carefully removed from the interior aspect of skulls and surfaces of brains with Dumont #5 forceps (Fine Science Tools). Brains were separated into sections (neocortex, hippocampus, striatum, cerebellum, brainstem) and sections from each group ( $n = 3$  mice per group) were pooled. Brain tissue was homogenized with three gentle strokes in 10 ml FACS buffer in a 15 ml Tenbroeck Homogenizer (Lowell) then gently pressed through 70  $\mu$ m nylon mesh cell strainers with sterile plastic plungers (BD Biosciences) to yield a single cell suspension. Cells were then centrifuged at 280g at 4 °C for 10 min, the supernatant removed and cells re-suspended in ice-cold FACS buffer. Myelin was removed using AutoMACS and myelin removal beads (Miltenyi Biotech). Cells were stained for extracellular markers with antibodies to CD11b conjugated to Alexa 780, CD45 conjugated to APC, CD4 conjugated to PE and MHC-II conjugated to Alexa 700 (eBioscience). All cells were fixed in 1% PFA in 0.1 M pH 7.4 PBS. Fluorescence data were collected with a CyAn ADP High-Performance Flow Cytometer (Dako) then analysed using Flowjo software. To obtain equivalent and accurate cell counts, cells were gated first using the LIVE/DEAD Fixable Dead Cell Stain Kit as per the manufacturer's instructions (Invitrogen), pulse width versus area to select singlet cells, forward scatter versus side scatter to eliminate debris, then by appropriate markers for cell type (for example, CD11b for myeloid-derived cells, or CD3 for T cells). Experiments were repeated twice with brains from  $n = 3$  mice each group (total of  $n = 6$  mice at 12 weeks after bone marrow transplantation).

**Floating section immunohistochemistry.** Free-floating sections were incubated with 10% normal serum (either goat or chicken) for 1 h at room temperature in PBS containing 0.1% Triton X-100 (Sigma), followed by incubation with appropriate dilutions of primary antibodies (anti-CD11b (eBioscience), 1:100; anti-MECP2 1:200 (Cell Signaling Technology); anti-cleaved caspase-3 (Cell Signaling Technology) 1:300; anti-GFP (Abcam) 1:2,000; anti-GFAP (Abcam) 1:1,000; anti-IGF-1 (R&D Systems) 1:100; anti-Lamp-1 (Abcam) 1:1000) for

24–48 h at 4 °C in the same buffer but without serum. Sections were then washed for 10 min three times at room temperature in 0.1% Triton X-100 (Sigma) in PBS followed by incubation with Alexa-fluor 488, 546, 594, 633 or 647 chicken/goat anti-mouse/rat/goat/chicken IgG antibodies (1:1,000, Invitrogen) for 1 h at room temperature. After 30 s in 1:20,000 DAPI reagent, sections were washed again with 0.1% Triton X-100 in 0.1 M PBS (3  $\times$  10 min) and mounted with Aqua-Mount (Lerner Laboratories) under coverslips.

**Analysis of Nissl staining.** Three mice were analysed per group (wild type, *Mecp2*<sup>–/y</sup>, wild type  $\rightarrow$  wild type, wild type  $\rightarrow$  *Mecp2*<sup>–/y</sup>). Three coronal brain sections containing hippocampus were selected per mouse. Sections were incubated in 0.1% Triton X-100 (Sigma) in PBS for 10 min at 25 °C. Two washes were performed in PBS for 5 min each at 25 °C. Sections were then incubated in 1:40 Neurotrace Fluorescent Nissl Stain 530/615 Red (Molecular Probes), diluted in PBS, for 20 min at 25 °C. Sections were again incubated in 0.1% Triton X-100 (Sigma) in PBS for 10 min at 25 °C, and two washes were performed in PBS for 5 min each at 25 °C. After a final wash, sections were mounted on glass slides using Aqua-Mount (Thermo Scientific), coverslipped and visualized by confocal microscopy. Each hippocampal CA1 region (two per tissue section) was imaged, and quantified using the free hand tool in ImageJ. A Wacom computer graphics tablet was used to outline the somas of CA1 neurons, and the 'Measure and Label' function in ImageJ was used to quantify the area of each soma. After quantification, 100 soma measurements were randomly selected per mouse. This resulted in data sets of 300 soma measurements per group. GraphPad Prism was used to calculate two-way ANOVAs with Bonferroni *post-hoc* test.

**Phagocytosis assay and analysis.** To determine the phagocytic ability of microglial cells, we chose apoptotic neural cells as targets, because they would most closely approximate natural targets *in vivo*. Accordingly, dissociated neural progenitor cells were treated with ultraviolet light for 15 min, stained with 5(6)-TAMRA, succinimidyl ester (Invitrogen), and washed thoroughly with cold PBS before they were used for incubation with cell culture and fed on to wild-type or *Mecp2*<sup>–/y</sup> microglial cells. After 2 or 5 hours at 37 °C and 5% CO<sub>2</sub>, cells were removed, then washed and fixed with 4% PFA O/N. Coverslips were blocked in 10% chicken serum in 0.1 M PBS containing 0.3% Triton X-100 and 0.5% BSA, followed by incubation with anti-CD11b (eBioscience, 1:100) for 1 h at room temperature. Coverslips were then washed for 10 min three times at room temperature in 0.1% Triton X-100 (Sigma) in PBS, followed by incubation with Alexa-fluor 488 chicken anti-rat IgG antibodies (1:1,000, Invitrogen) for 1 h at room temperature. Coverslips were washed again with 0.1% Triton X-100 in 0.1 M PBS (10 min, three times) and mounted with Aqua-Mount (Lerner Laboratories) on slides. Slides were analysed using confocal microscopy and ImageJ software.

**Whole-animal plethysmography.** Plethysmography was performed during the dark phase of the diurnal cycle from 9:30 to 12:30 to assess normal waking respiratory parameters. Respiratory parameters were assessed by whole-animal plethysmography in unrestrained C57Bl/6J (wild-type), *Mecp2*<sup>–/y</sup>, wild-type  $\rightarrow$  *Mecp2*<sup>–/y</sup>, *Mecp2*<sup>–/y</sup>  $\rightarrow$  *Mecp2*<sup>+/y</sup>, *Lysm*<sup>cre</sup>, *Mecp2*<sup>lox-stop/y</sup>*Lysm*<sup>cre</sup> and *Mecp2*<sup>lox-stop/y</sup> mice. Animals were placed individually into 1,000 cm<sup>3</sup> Plexiglas chambers (Buxco) and allowed 15 min to acclimate. The chamber was continuously flushed with dry room-temperature air ( $24 \pm 0.5$  °C) delivered by three computer-driven mass-flow regulators connected to pure O<sub>2</sub>, N<sub>2</sub> and CO<sub>2</sub> (total flow: 1 l min<sup>–1</sup>). The flow signal was recorded and analysed using EMKA IOX 2.7 (EMKA Technologies) and Spike 5.21 (CED) software. Breathing parameters were calculated from a calibrated flow signal derived from a differential pressure sensor connected to the plethysmography chamber using the equation of Drorbaugh and Fenn<sup>30</sup>. Real-time chamber conditions (temperature and humidity) and atmospheric pressure were continuously measured and the calculation of tidal volume was automatically adjusted to account for changes in these variables. Inspiration and expiration were detected using a pressure transducer calibrated before each experiment by injecting 1 ml of air. The pressure signal was amplified, digitized and recorded using IOX software (EMKA Technologies) and Spike2 software (Cambridge Electronic Design), concurrently. Animals were left in chambers until sufficient data representing periods of quiescent breathing were collected to allow analysis. Animals were then returned to home cages.

**Analysis of plethysmography.** For apnoea scoring, plethysmography traces were hand-scored for apnoeas using the following criteria: expiratory time,  $T_E > 1$  s measured exclusively during periods free of grooming, sniffing or locomotor behaviour, such that false positives were excluded.

**Interbreath irregularity.** Interbreath variability was derived from the absolute value of  $((T_{\text{total},n} - T_{\text{total},n+1})/T_{\text{total},n+1})$ , where  $T_{\text{total},n}$  is the duration of the breathing cycle during the  $n$ th breath. In each mouse this variable was determined over 100 consecutive breaths during periods when the animals were quiescent (no locomotion or sniffing) and the average value produced a single score per mouse.

This score was averaged for each group of mice and expressed as a percentage ( $\pm$  s.e.m.).

**Open-field activity.** Behavioural testing in a novel open field was also performed during the dark phase of the diurnal cycle from 21:30 to 23:30 in the vivarium, such that measured locomotion would most closely correspond to normal waking activity. Animals were brought to the testing location within the vivarium and allowed to acclimate for 30 min before beginning the assay. Open field behaviour was measured using a Versamax activity monitor (AccuScan Instruments). Each mouse was placed initially in the centre of the monitoring cage, and allowed to roam freely for 20 min before being returned to the home cage. Animal activity levels were examined using Versamax software.

**Accelerating rotarod.** Motor coordination and balance were assessed on an Economex accelerating rotarod (Columbus Instruments) that had the capacity to test four mice simultaneously. The testing procedure consisted of two training phases and a testing phase: stationary training on a non-rotating rod, constant-speed training on a rod rotating at a speed of 2.0 r.p.m. and testing on an accelerating rotarod (acceleration = 0.1 r.p.m.). Latency to fall from the rotarod was recorded. For stationary training, the mouse was placed on the non-rotating rod facing the back side of the apparatus. The mouse was required to remain on the rod for 60 s before it was removed from the apparatus. If the mouse fell before the 60 s cutoff per trial, the animal was placed back onto the rod. This procedure was repeated for two trials with no inter-trial interval. For constant-speed training, the rod was adjusted to spin at a constant speed of 2.0 r.p.m. The mouse was placed on the rod facing the back side of the apparatus, and was required to remain on the rotating rod for 60 s before it was removed from the apparatus. If the mouse fell before the 60 s cutoff per trial, the animal was placed back onto the rod. This procedure was repeated for a total of two trials with no inter-trial interval. For accelerating rotarod testing, the rod was adjusted to spin at a constant speed of 2.0 r.p.m. The acceleration (acceleration = 0.1 r.p.m.) was started and the latency to fall recorded.

**Microglia primary culture.** Mouse mixed glia cultures were prepared from P2 mouse neonates as follows. Brains were excised and placed in ice-cold HBSS. Neocortical tissue was removed, meninges discarded and minced with forceps. Minced tissue was then incubated in 2 mg ml<sup>-1</sup> papain (Sigma) in HBSS at 37 °C for 30 min. After adding cold heat-inactivated fetal bovine serum, and DNase, the tissue was washed twice with cold HBSS by re-suspending the tissue and then pelleting in a 4 °C centrifuge. To obtain a single-cell suspension, the tissue was triturated gently by pipetting through a 5 ml serological pipette 20 times and then filtered through a 70 µm nylon filter (BD Biosciences). Mixed glial cells were cultured with culture medium consisting of DMEM/F12 with 10% fetal bovine serum (Invitrogen), GlutaMAX (Invitrogen) and 1% penicillin/streptomycin in a 5% CO<sub>2</sub>/37 °C incubator, changing medium after 7 days, for a total of 14 days. To

obtain microglia, flasks were shaken at 120 r.p.m. for 12 h at 37 °C. Supernatants were collected, spun at 300g for 7 min, and then washed twice by re-suspending the cells in fresh medium and centrifuging at 25 °C. Cell pellets were re-suspended in culture medium and seeded at  $5 \times 10^5$  cells per millilitre onto 24-well inserts ( $5 \times 10^5$  cells per well). Microglia were allowed to rest for 48 h before MHCII and cytokine expression assays.

**Bone-marrow-derived macrophage culture.** Macrophage cultures were prepared from bone marrow as follows: bone marrow was isolated as above. Five hundred thousand cells were pipetted into each well of a 24-well plate in 1 ml culture medium consisting of DMEM/F12 with 10% fetal bovine serum (Invitrogen), 10 ng ml<sup>-1</sup> rMCSF (eBioscience) and 1% penicillin/streptomycin in a 5% CO<sub>2</sub>/37 °C incubator, changing 500 µl medium after 3 days and 6 days. At 9 days, mature macrophages were washed three times with 37 °C 0.1 M PBS to remove all non-adherent cells, then placed in 1 ml culture medium consisting of DMEM/F12 with 10% fetal bovine serum (Invitrogen) and 1% penicillin/streptomycin. Cells were allowed to rest for 6 h before further treatment.

**Enzyme-linked immunosorbent assay of bone-marrow-derived macrophage supernatants.** Bone-marrow-derived macrophages were prepared as above. Cells were treated in triplicate wells with cytokines (50 ng ml<sup>-1</sup> rIL4 or rIFNγ), or plain medium (control) and placed in a 5% CO<sub>2</sub>/37 °C incubator for 72 h. If cells were subsequently treated with lipopolysaccharide (LPS; 100 ng ml<sup>-1</sup>), they were removed from the incubator at 48 h, LPS was added and cells replaced for the final 24 h of incubation. At 72 h, 1 ml supernatant was removed from each well and frozen immediately at -80 °C. Samples were thawed simultaneously, an aliquot of each diluted 1:5 in manufacturer's diluent and enzyme-linked immunosorbent assay was performed with triplicate samples, according to the manufacturer's instructions (R&D).

**TUNEL.** Sections (20 µm) were sliced by cryostat and assayed the same day. Sections were permeabilized for 1 h using neuropore reagent (Trevigen) then labelled using the *In situ* Cell Death Detection Kit, Fluorescein (Roche) according to the manufacturer's instructions (using floating sections). Sections were washed  $3 \times 10$  min in 0.1 M PBS, incubated for 30 s in 1:20,000 DAPI reagent, washed  $1 \times 10$  min in 0.1 M PBS, mounted on slides with Aqua-Mount (Lerner) and coverslipped. Images were taken using a Zeiss LSM 700 confocal microscope. Six total slices from the cerebellum of three mice were imaged per group. Five images were taken per slice, and a randomly selected 100 µm  $\times$  100 µm section of the granular layer was sampled for each image. The area of staining was measured for each field. The staining per field was used to calculate statistics by two-way ANOVA with Bonferroni *post hoc* test using GraphPad Prism.

30. Drorbaugh, J. E. & Fenn, W. O. A barometric method for measuring ventilation in newborn infants. *Pediatrics* **16**, 81–87 (1955).

# Notch-dependent VEGFR3 upregulation allows angiogenesis without VEGF–VEGFR2 signalling

Rui Benedito<sup>1,2</sup>, Susana F. Rocha<sup>1,2</sup>, Marina Woeste<sup>1,2</sup>, Martin Zamykal<sup>1,2</sup>, Freddy Radtke<sup>3</sup>, Oriol Casanovas<sup>4</sup>, Antonio Duarte<sup>5</sup>, Bronislaw Pytowski<sup>6</sup> & Ralf H. Adams<sup>1,2</sup>

Developing tissues and growing tumours produce vascular endothelial growth factors (VEGFs), leading to the activation of the corresponding receptors in endothelial cells. The resultant angiogenic expansion of the local vasculature can promote physiological and pathological growth processes<sup>1</sup>. Previous work has uncovered that the VEGF and Notch pathways are tightly linked. Signalling triggered by VEGF-A (also known as VEGF) has been shown to induce expression of the Notch ligand DLL4 in angiogenic vessels and, most prominently, in the tip of endothelial sprouts<sup>2,3</sup>. DLL4 activates Notch in adjacent cells, which suppresses the expression of VEGF receptors and thereby restrains endothelial sprouting and proliferation<sup>2,4–6</sup>. Here we show, by using inducible loss-of-function genetics in combination with inhibitors *in vivo*, that DLL4 protein expression in retinal tip cells is only weakly modulated by VEGFR2 signalling. Surprisingly, Notch inhibition also had no significant impact on VEGFR2 expression and induced deregulated endothelial sprouting and proliferation even in the absence of VEGFR2, which is the most important VEGF-A receptor and is considered to be indispensable for these processes. By contrast, VEGFR3, the main receptor for VEGF-C, was strongly modulated by Notch. VEGFR3 kinase-activity inhibitors but not ligand-blocking antibodies suppressed the sprouting of endothelial cells that had low Notch signalling activity. Our results establish that VEGFR2 and VEGFR3 are regulated in a highly differential manner by Notch. We propose that successful anti-angiogenic targeting of these receptors and their ligands will strongly depend on the status of endothelial Notch signalling.

Angiogenesis is prominently controlled by the activity of Notch in the endothelium of developing mice, zebrafish embryos and tumour models<sup>6–9</sup>. Although the precise mechanisms linking the Notch and VEGF pathways remain incompletely understood, the activation of DLL4–Notch by VEGFR2 and the repression of VEGFR2 expression downstream of Notch activation are seen as two crucial processes regulating endothelial sprouting and angiogenesis<sup>1,4,5,10,11</sup>. To study the role of VEGFR2 *in vivo* and circumvent the embryonic lethality of a global knockout of *Vegfr2* (also known as *Flk1* and *Kdr*) in mice, we generated inducible endothelial-cell-specific loss-of-function mutants (*Vegfr2*<sup>iAEC</sup>) by combining *loxP*-flanked (floxed) *Vegfr2* alleles (*Vegfr2*<sup>flxed/flxed</sup>)<sup>12</sup> and tamoxifen-inducible Cre (*CreERT2*) transgenes<sup>13,14</sup>. Following tamoxifen administration from postnatal day 3 (P3) to P5, analysis of the retinal vasculature at P7 confirmed that endothelial sprouting and vessel density were strongly reduced in *Vegfr2*<sup>iAEC</sup> mutants (Fig. 1a, b).

Surprisingly, blocking Notch activation (for 48 h before analysis) with the  $\gamma$ -secretase inhibitor DAPT (*N*-[*N*-(3,5-difluorophenacetyl)-L-alanyl]-*S*-phenylglycine *t*-butyl ester) strongly enhanced angiogenesis even in these mutant animals (Fig. 1a, b and Supplementary Fig. 2a, b). Because Notch is not the only substrate of  $\gamma$ -secretase and because

DAPT acts on all cell types, we combined *Vegfr2*<sup>flxed/flxed</sup> and *Dll4*<sup>flxed/flxed</sup> mice<sup>15</sup> to generate endothelial-cell-specific double mutants. Deletion of the *Dll4* gene between P1 and P3 prominently increased sprout formation and vascular density (Fig. 1c), the latter of which reflects the known suppression of endothelial cell proliferation by Notch<sup>3,6,7</sup>. Confirming the DAPT data, this increase in angiogenesis was not abolished by also targeting *Vegfr2*, even though both gene products were no longer detectable at the protein level (Fig. 1c and Supplementary Fig. 2c, d). The outgrowth and density of the *Dll4*<sup>iAEC</sup>/*Vegfr2*<sup>iAEC</sup> vasculature were slightly reduced compared with those of *Dll4*<sup>iAEC</sup> mutants, whereas the numbers of sprouts and filopodia at the angiogenic front (that is, the distal region of the growing vasculature) were similar (Fig. 1e and Supplementary Fig. 2c, f). We also compared angiogenesis in the retinas of mice with different combinations of floxed *Notch1* and *Vegfr2* alleles and did not observe any significant differences in vascular density or sprouting between *Notch1*<sup>iAEC</sup> and *Notch1*<sup>iAEC</sup>/*Vegfr2*<sup>iAEC</sup> retinas (Supplementary Fig. 2e, f).

The *Rbpj* gene encodes the transcription factor RBP-J $\kappa$ , which binds to the intracellular domain of activated Notch (NICD) and is indispensable for Notch signalling<sup>16</sup>. The phenotype of inducible endothelial-cell-specific *Rbpj*<sup>iAEC</sup> mutants manifests itself only after P4, which enabled us to study the effects of VEGFR2 and VEGF-A inhibition in this model from P4 to P7 (Fig. 1d). Like the targeting of the *Vegfr2* gene, the administration of VEGF-A- or VEGFR2-blocking antibodies efficiently reduced endothelial cell sprouting and vessel density in the retinas of control mice but did not significantly impair the deregulated angiogenic growth in *Rbpj*<sup>iAEC</sup> mutants (Fig. 1d, e and Supplementary Fig. 3). This finding further supported the idea that angiogenesis in Notch-deficient vessels can occur in the absence of VEGF-A or VEGFR2 activity. Moreover, despite the high vessel density and extensive sprouting in the *Rbpj*<sup>iAEC</sup> and *Dll4*<sup>iAEC</sup> vasculature, vessel outgrowth towards the VEGF-A-rich (avascular) retinal periphery was strongly reduced (Fig. 1d, e). Thus, in contrast to the existing models of the crosstalk between VEGF and Notch, these results suggest that endothelial responses to the tissue-derived VEGF-A gradient are not enhanced by the loss of Notch activity.

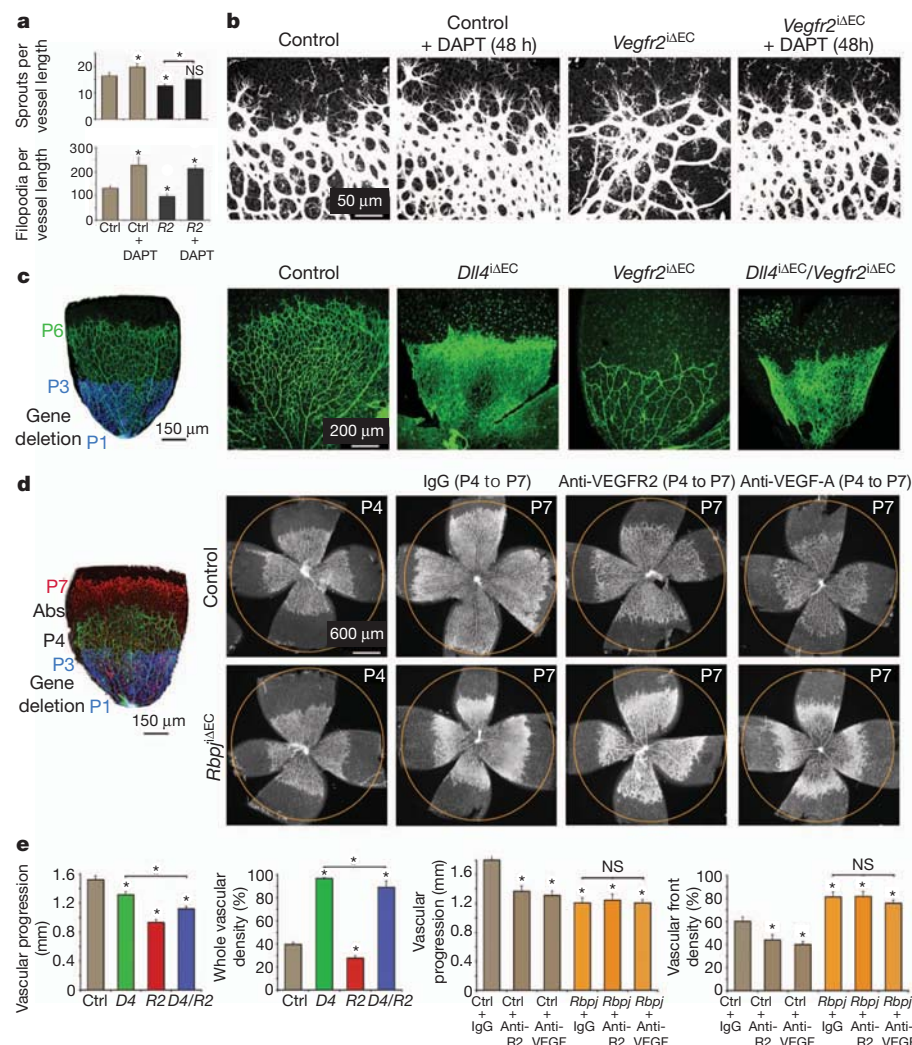
Previous studies have reported that VEGF-A–VEGFR2 signalling is important for the upregulation of *Dll4* expression in sprouting endothelial tip cells<sup>2,3</sup>, which, in turn, is thought to activate Notch signalling in adjacent (stalk) cells and thereby suppress sprouting behaviour. However, the strong effects of *Dll4* or *Notch* inactivation in the *Vegfr2*<sup>iAEC</sup> background (Fig. 1a–e) suggested that a substantial level of endothelial Notch signalling was maintained without VEGFR2 function. Indeed, control and *Vegfr2*<sup>iAEC</sup> sprouting tip cells and arteries displayed similar anti-DLL4 antibody staining (Fig. 2a and Supplementary Fig. 4), and weaker protein signals were only seen within the mutant capillary network. *Dll4* and *Hey1* messenger RNA levels were also only partially reduced in *Vegfr2*<sup>iAEC</sup> lungs, a tissue that was

<sup>1</sup>Max Planck Institute for Molecular Biomedicine, Department of Tissue Morphogenesis, D-48149 Münster, Germany. <sup>2</sup>University of Münster, Faculty of Medicine, D-48149 Münster, Germany. <sup>3</sup>Ecole Polytechnique Fédérale de Lausanne (EPFL), Swiss Institute for Experimental Cancer Research (ISREC), CH-1015 Lausanne, Switzerland. <sup>4</sup>Translational Research Laboratory, Catalan Institute of Oncology, IDIBELL, 08907 L'Hospitalet de Llobregat, Spain. <sup>5</sup>The Interdisciplinary Centre of Research in Animal Health (CIISA), Faculty of Veterinary Medicine, Technical University of Lisbon, 1300-474 Lisbon, Portugal. <sup>6</sup>ImClone Systems, 180 Varick Street, New York, New York 10014, USA.



# Figure 1 | Notch inhibition promotes angiogenesis independently of VEGFR2.

**a, b,** Quantification of sprouting defects and representative confocal images of the isolectin-B4-stained angiogenic front of *Vegfr2*<sup>ΔEC</sup> mutants, as well as the effect of DAPT treatment (see also Supplementary Fig. 2a, b). Ctrl, control; R2, *Vegfr2*<sup>ΔEC</sup>. **c,** Induction of endothelial-cell-specific *Dll4* or *Vegfr2* deletion from P1 to P3 and analysis at P6. Note the strongly enhanced angiogenesis in *Dll4*<sup>ΔEC</sup> and *Dll4*<sup>ΔEC</sup>/*Vegfr2*<sup>ΔEC</sup> mutants compared with control and *Vegfr2*<sup>ΔEC</sup> retinas. **d,** Isolectin-B4-stained control and *Rbpj*<sup>ΔEC</sup> P4 and P7 retinas (tamoxifen administration from P1 to P3) treated with control IgG, or anti-VEGFR2 or anti-VEGF-A antibodies (Abs) from P4 to P7. The orange circles facilitate the comparison of vascular progression. **e,** Quantification of vascular parameters of retinas in **c** and **d**. See Supplementary Figs 2 and 3 for additional data. *D4*, *Dll4*<sup>ΔEC</sup>; *D4/R2*, *Dll4*<sup>ΔEC</sup>/*Vegfr2*<sup>ΔEC</sup>; *Rbpj*, *Rbpj*<sup>ΔEC</sup>. **a, e,** Data are presented as mean + s.e.m.; *n* ≥ 12. \*, *P* < 0.05 compared with the left-most histogram bar unless otherwise indicated; NS, not significant.



chosen for analysis because of its particularly high endothelial cell content (Fig. 2c). The level of NICD protein was lower in *Vegfr2*<sup>ΔEC</sup> lungs, but NICD was still readily detectable by immunoblotting (Fig. 2c). By contrast, VEGFR3, which has previously been shown to be positively regulated by VEGFR2 signalling<sup>5</sup>, showed the expected decrease in expression in the *Vegfr2*<sup>ΔEC</sup> endothelium (Fig. 2b, c and Supplementary Fig. 4c). These results indicate that substantial DLL4 expression is maintained in the absence of VEGFR2, while the VEGFR3 protein levels are strongly reduced.

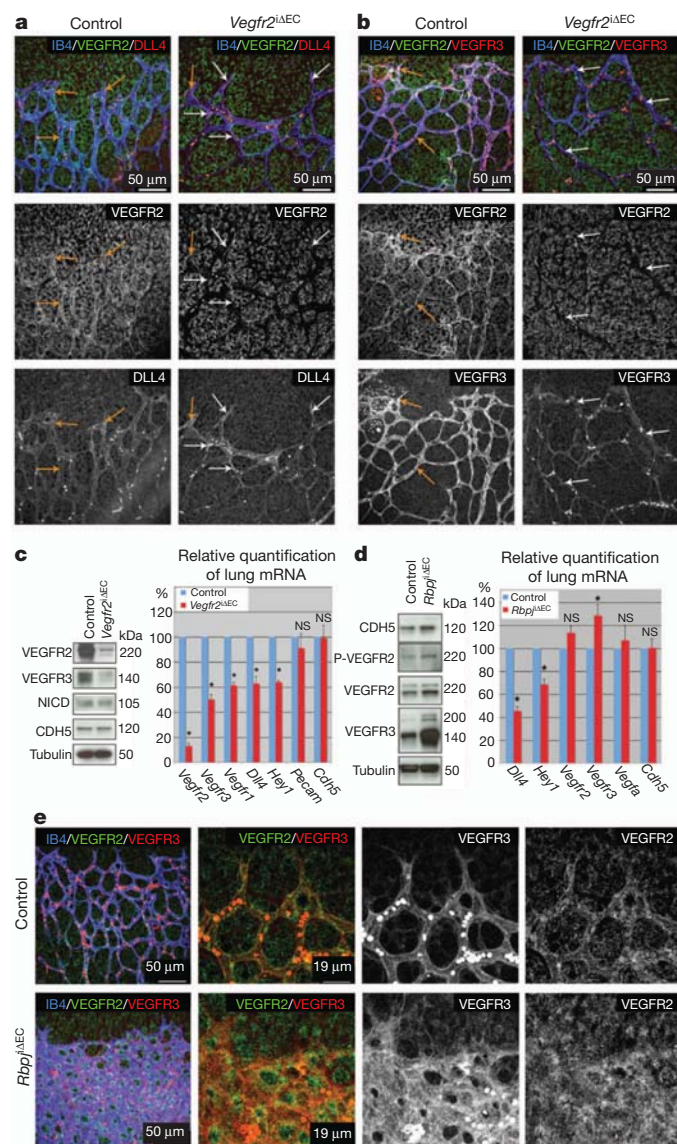
Active Notch is widely seen as a potent negative regulator of VEGFR2 and VEGFR3 expression<sup>4,11</sup>, but published reports provide an inconsistent picture. For example, overexpression of DLL4 or NICD in cultured human umbilical vein endothelial cells (HUVECs) was shown to repress *VEGFR2* transcription<sup>17,18</sup>. By contrast, DAPT treatment of HUVECs did not lead to significant changes in VEGFR2 protein or *VEGFR2* transcript levels, despite increased proliferation and sprouting<sup>17,19</sup> (Supplementary Fig. 5a). In the mouse retina, both *Vegfr2* and *Vegfr3* mRNAs have been found to be only slightly upregulated in *Dll4*<sup>+/-</sup> heterozygotes or after DAPT administration<sup>2,5</sup>. In zebrafish, *vegfr3* (also known as *flt4*) mRNA expression has been shown to be increased in *rbpj* morphants but not in *dll4* morphants<sup>6,20</sup>. Our own data indicate the selective regulation of VEGFR3 but not VEGFR2 downstream of Notch. In arteries, in which Notch is more active, DLL4 was strongly expressed; however, the endogenous levels of the VEGFR3 protein were low. By contrast, VEGFR2 was present in arteries and veins (Supplementary Fig. 5b–d).

Furthermore, we found a strong increase in the VEGFR3 protein but not the VEGFR2 protein in the *Rbpj*<sup>ΔEC</sup> angiogenic front (Fig. 2e).

Upregulation of the VEGFR3 protein was also observed in lung lysates and sections from *Rbpj*<sup>ΔEC</sup> and *Dll4*<sup>ΔEC</sup> mutants, whereas the VEGFR2 level was not significantly altered (Figs 2d and 3c and Supplementary Fig. 6). Interestingly, this difference was not mirrored at the mRNA level: the levels of *Vegfr2* and *Vegfr3* mRNA were not substantially increased, which hints at post-transcriptional regulation of VEGFR3 levels by Notch (Fig. 2d). The differential regulation of VEGFR protein levels was also mirrored by the activity levels of these receptors. The level of phosphorylated VEGFR3 (phospho-VEGFR3) was strongly increased in *Rbpj*<sup>ΔEC</sup> lung lysates, whereas the increase in phospho-VEGFR2 was more modest and mirrored the increase in vascular endothelial cadherin (CDH5) (which reflects the number of endothelial cells) in the *Rbpj*<sup>ΔEC</sup> mutant model (Figs 2d and 3a–c). VEGFR2 and Notch play opposing roles in the regulation of VEGFR3 levels. Their actions seem to be independent because the combined deletion of *Dll4* and *Vegfr2* (*Dll4*<sup>ΔEC</sup>/*Vegfr2*<sup>ΔEC</sup>) or the administration of DAPT to *Vegfr2*<sup>ΔEC</sup> pups led to a significant increase in VEGFR3 protein compared with control or *Vegfr2*<sup>ΔEC</sup> mutant pups (Fig. 3c, d; see also Fig. 2b).

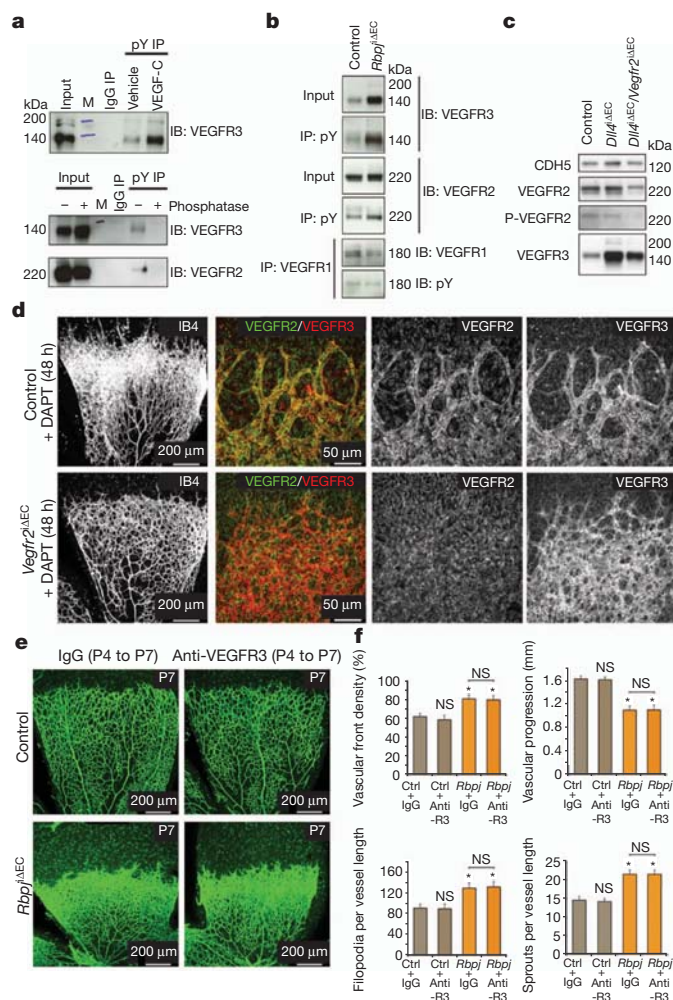
To investigate the functional consequences of high levels and activity of VEGFR3 protein in endothelial cells with low or compromised Notch signalling, we first used a monoclonal antibody that specifically blocks ligand binding to VEGFR3 (mF4-31C1). Daily injection from P4 to P7 with 50 mg kg day<sup>-1</sup> of mF4-31C1 did not impair vessel growth in control or *Rbpj*<sup>ΔEC</sup> mutants (Fig. 3e, f), suggesting that the ligand-dependent activation of VEGFR3 is not important for developmental angiogenesis, which is consistent with previous reports<sup>21,22</sup>. We then carried out short-term administration (24 h) of DAPT and the kinase inhibitor MAZ51. MAZ51 potently inhibits VEGF-C- or





**Figure 2 | VEGFR2 strongly regulates VEGFR3 protein levels but not DLL4 at the angiogenic front.** **a, b**, Whole-mount triple immunofluorescence for DLL4 or VEGFR3 (red), VEGFR2 (green) and isolectin B4 (IB4; blue) of P6 control and *Vegfr2*<sup>ΔEC</sup> retinas. The deletion of *Vegfr2* for 5 days strongly compromised angiogenesis and the expression of VEGFR3 but not of DLL4 in tip endothelial cells. The arrows indicate endothelial cells expressing VEGFR2 (orange) and devoid of VEGFR2 (white). **c, d**, Immunoblotting (left) and quantitative PCR analysis with reverse transcription (qRT-PCR; right) of *Vegfr2*<sup>ΔEC</sup> and *Rbpj*<sup>ΔEC</sup> P6 mouse lungs for the indicated proteins and transcripts. The CDH5 protein levels reflect endothelial cell content. Data are presented as mean ± s.e.m.; *n* = 4. \*, *P* < 0.001 compared with control group data (blue bars) unless otherwise indicated; NS, not significant. P-VEGFR2, phospho-VEGFR2. **e**, Whole-mount triple immunofluorescence for VEGFR3 (red), VEGFR2 (green) and isolectin B4 (blue), showing upregulated VEGFR3 but not VEGFR2 protein in *Rbpj*<sup>ΔEC</sup> vessels. **a, b, e**, Autofluorescent red blood cells are visible as round spots inside the vessels.

VEGF-D-induced activation of VEGFR3 but only weakly impairs VEGFR2 activation by VEGF-A<sup>23</sup>. In contrast to the VEGFR3-blocking antibody, MAZ51 strongly inhibited both VEGF-C-dependent and VEGF-C-independent VEGFR3 phosphorylation in cultured cells (Supplementary Fig. 7). Although MAZ51 alone only weakly affected the number of filopodia and sprouts at the angiogenic front of the control retinal vasculature, this inhibitor strongly suppressed the enhanced sprouting caused by DAPT (Fig. 4a, b). MAZ51 also reduced the phosphorylation of VEGFR3 in *Rbpj*<sup>ΔEC</sup> lung lysates (Fig. 4c).



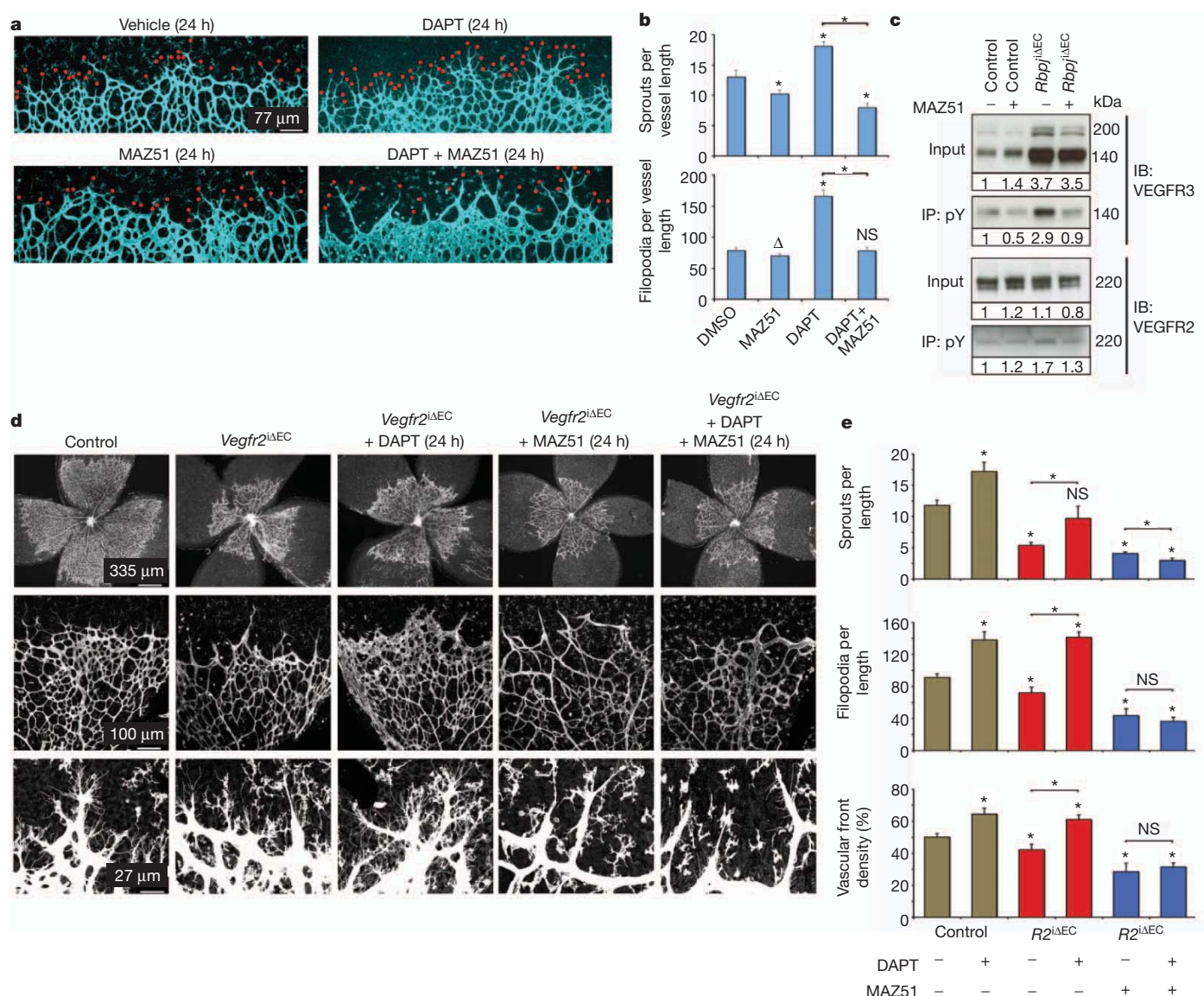
**Figure 3 | Notch regulates VEGFR3 activity independently of VEGFR2.**

**a**, Controls for VEGFR3 and VEGFR2 phosphotyrosine (pY) immunoprecipitation (IP) from P6 lungs. Phospho-VEGFR3 is present at increased levels in VEGF-C-injected P6 pups. The specificity was confirmed by lambda protein phosphatase pre-incubation. IB, immunoblotting antibody; M, molecular weight markers. **b**, Phosphotyrosine immunoprecipitation from lung lysates, showing strongly increased phospho-VEGFR3 in *Rbpj*<sup>ΔEC</sup> mutants. Smaller changes were observed for phospho-VEGFR2, total VEGFR1 and phospho-VEGFR1. **c**, Immunoblot analysis, showing strongly increased VEGFR3 protein levels in *Dll4*<sup>ΔEC</sup> and *Dll4*<sup>ΔEC</sup>/*Vegfr2*<sup>ΔEC</sup> P6 lung lysates, which was only partly attenuated in the absence of endothelial VEGFR2. **d**, Isolectin B4 combined with VEGFR2 (green) and VEGFR3 (red) staining of control and *Vegfr2*<sup>ΔEC</sup> retinas. DAPT administration for 48 h before dissection restored high VEGFR3 protein levels at the *Vegfr2*<sup>ΔEC</sup> angiogenic front (compare with Fig. 2b). **e**, Confocal images of isolectin-B4-stained P7 control and *Rbpj*<sup>ΔEC</sup> retinas (with tamoxifen administration from P1 to P3) injected with control IgG or anti-VEGFR3 antibody from P4 to P7. **f**, Quantification of vascular parameters in control and *Rbpj*<sup>ΔEC</sup> mice treated with VEGFR3-blocking antibody (anti-R3) or IgG, as indicated. Data are presented as mean ± s.e.m.; *n* ≥ 12. \*, *P* < 0.05 compared with control group data shown to the left of each graph unless otherwise indicated; NS, not significant.

These results point to a crucial role for VEGFR3 in the regulation of sprouting in endothelial cells with low Notch signalling activity, which is also in agreement with previous reports using morpholino targeting of *vegfr3* in zebrafish<sup>6,20</sup>.

By contrast, the DAPT-induced increase in endothelial cell density at the angiogenic front was not significantly altered by the addition of MAZ51 (Fig. 4a and Supplementary Fig. 8). Providing further evidence for the important role of VEGFR3 in Notch-controlled angiogenesis, the injection of *Vegfr2*<sup>ΔEC</sup> pups with DAPT and MAZ51 (at P5 for 24 h) led to a striking decrease in the number of sprouts and filopodia





**Figure 4 | Inhibition of VEGFR3 kinase activity suppresses Notch-regulated sprouting.** **a, b,** Retina whole-mount immunostaining for islectin B4 (blue) after 24 h treatment (P5 to P6) with the indicated inhibitors. MAZ51 efficiently blocked sprouting (red dots) of DAPT-treated endothelial cells without having an overt effect on vessel density (see also Supplementary Fig. 8). Vehicle, dimethylsulphoxide (DMSO). **c,** MAZ51 had strong effects on VEGFR3 but not VEGFR2 phosphorylation *in vivo*. Lungs from pups with the indicated genotypes and treatments were subjected to phosphotyrosine

and the vascular density compared with DAPT-treated *Vegfr2*<sup>ΔEC</sup> littermates (Fig. 4d, e). Moreover, administration of AAL-993, a kinase inhibitor that blocks all VEGFRs, led to a significant decrease in vascular density and endothelial sprouting in DAPT-treated retinas (Supplementary Fig. 9). These results indicate that the inhibition of Notch signalling allows vessel growth in the absence of VEGF-A–VEGFR2 signalling, and this vessel growth requires the kinase activity of VEGFR3 but not the binding of VEGF-C. This is consistent with recent reports proposing ligand-independent VEGFR3 activation<sup>24,25</sup>.

The initially simplistic models describing the molecular crosstalk between the VEGF and Notch pathways need to be revised. It has been reported that endothelial-cell-specific deletion of the *Vegfr3* gene postnatally in mice leads to enhanced angiogenesis<sup>24</sup>. The authors of this report proposed that VEGFR3 has ‘active’ (ligand-dependent and pro-angiogenic) and ‘passive’ (ligand-independent and anti-angiogenic) signalling modes, among which the passive mode is mediated by

immunoprecipitation followed by VEGFR3 and VEGFR2 immunoblotting. The numbers below the blots indicate the relative quantification of the signals. **d, e,** Islectin-B4-stained control and *Vegfr2*<sup>ΔEC</sup> P5 retinas after 24 h treatment with the indicated chemical inhibitors. Inhibition of the kinase activity of VEGFR3 strongly impaired angiogenesis in DAPT-treated *Vegfr2*<sup>ΔEC</sup> mutants. **b, e,** Data are presented as mean + s.e.m.;  $n \geq 12$ . \*,  $P < 0.05$ ;  $\Delta$ ,  $P = 0.1$ ; NS, not significant; compared with control group data shown to the left of each graph unless otherwise indicated.

Notch activation. Our own data indicate that VEGFR3 activity is pro-angiogenic in endothelial cells with low or no Notch signalling activity. Both findings are not incompatible and might reflect a dynamic, Notch-controlled switching of VEGFR3 function in growing blood vessels.

Our results also fundamentally change the understanding of the crosstalk between Notch and VEGFR2 in angiogenesis (Supplementary Fig. 1). VEGFR2 is not essential for *Dll4* expression in tip cells, suggesting that other upstream regulators, such as cell–matrix signalling, have important roles<sup>26,27</sup>. Although VEGFR2 is not regulated or only weakly regulated by Notch, both Notch and VEGFR2 modulate VEGFR3 protein and activation levels independently and in opposite directions. In settings with low Notch activity, VEGFR3 upregulation allows strong, ligand-independent and highly deregulated angiogenesis even in the absence of VEGF–VEGFR2 signalling, which mimics aspects of growth-factor-independent (autocrine) cancer cell growth<sup>28</sup>. This mode of vessel growth might contribute to the excessive and



deregulated angiogenesis that is observed in tumours treated with Notch inhibitors<sup>9</sup>. Probing the status of vascular Notch or VEGFR3 activation might be relevant for patients who do not respond to anti-VEGF treatment, a complicating factor in the therapy of cancer and age-related macular degeneration<sup>29,30</sup>. This information could yield valuable clues for selecting suitable therapeutic strategies that could reduce the resistance to current anti-angiogenic therapies.

## METHODS SUMMARY

**Mutant mice, inducible genetic experiments and pharmacological inhibition.** For inducible and endothelial-cell-specific genetic loss-of-function experiments, we interbred *Vegfr2*<sup>flxed</sup> (ref. 12), *Dll4*<sup>flxed</sup> (ref. 15), *Notch1*<sup>flxed</sup> (see Methods) or *Rbpj*<sup>flxed</sup> (ref. 16) mice with transgenic mice expressing the tamoxifen-inducible recombinase CreERT2 under the control of the endothelial *Cdh5* (ref. 14) or *Pdgfrb* (ref. 13) promoter.

Cre activity and gene deletion were induced by intraperitoneal injections of 50 µg (P1–P3) or 100 µg (P3–P5) tamoxifen (Sigma, T5648; 2 mg ml<sup>-1</sup>). In some experiments, Notch signalling was inhibited in half of the pups by subcutaneous injection with 100 mg kg<sup>-1</sup> DAPT (Calbiochem). The kinase activity of VEGFR3 was inhibited with 10 mg kg<sup>-1</sup> MAZ51 (Sigma). For simultaneous inhibition of the kinase activity of all VEGFRs, we used 25 mg kg<sup>-1</sup> AAL-993 (Enzo Life Sciences). To block signalling with monoclonal antibodies, daily intraperitoneal injections with 50 mg kg<sup>-1</sup> anti-VEGFR2 (ImClone Systems, DC101), anti-VEGFR3 (ImClone Systems, mF4-31C1), anti-VEGF-A (Genentech, G6-31) or control mouse or rat IgG antibodies (Jackson ImmunoResearch) were given at P4, P5 and P6.

**Ethical review.** All animal experiments were performed in compliance with the relevant laws and institutional guidelines and were approved by local animal ethics committees.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 1 December 2010; accepted 26 January 2012.

Published online 18 March 2012.

- Lohela, M., Bry, M., Tammela, T. & Alitalo, K. VEGFs and receptors involved in angiogenesis versus lymphangiogenesis. *Curr. Opin. Cell Biol.* **21**, 154–165 (2009).
- Suchting, S. *et al.* The Notch ligand Delta-like 4 negatively regulates endothelial tip cell formation and vessel branching. *Proc. Natl Acad. Sci. USA* **104**, 3225–3230 (2007).
- Lobov, I. B. *et al.* Delta-like ligand 4 (Dll4) is induced by VEGF as a negative regulator of angiogenic sprouting. *Proc. Natl Acad. Sci. USA* **104**, 3219–3224 (2007).
- Phng, L. K. & Gerhardt, H. Angiogenesis: a team effort coordinated by Notch. *Dev. Cell* **16**, 196–208 (2009).
- Tammela, T. *et al.* Blocking VEGFR-3 suppresses angiogenic sprouting and vascular network formation. *Nature* **454**, 656–660 (2008).
- Siekmann, A. F. & Lawson, N. D. Notch signalling limits angiogenic cell behaviour in developing zebrafish arteries. *Nature* **445**, 781–784 (2007).
- Benedito, R. *et al.* The Notch ligands Dll4 and Jagged1 have opposing effects on angiogenesis. *Cell* **137**, 1124–1135 (2009).
- Hellström, M. *et al.* Dll4 signalling through Notch1 regulates formation of tip cells during angiogenesis. *Nature* **445**, 776–780 (2007).
- Noguera-Troise, I. *et al.* Blockade of Dll4 inhibits tumour growth by promoting non-productive angiogenesis. *Nature* **444**, 1032–1037 (2006).
- Siekmann, A. F., Covassin, L. & Lawson, N. D. Modulation of VEGF signalling output by the Notch pathway. *Bioessays* **30**, 303–313 (2008).
- Thurston, G. & Kitajewski, J. VEGF and Delta-Notch: interacting signalling pathways in tumour angiogenesis. *Br. J. Cancer* **99**, 1204–1209 (2008).
- Haigh, J. J. *et al.* Cortical and retinal defects caused by dosage-dependent reductions in VEGF-A paracrine signaling. *Dev. Biol.* **262**, 225–241 (2003).

- Claxton, S. *et al.* Efficient, inducible Cre-recombinase activation in vascular endothelium. *Genesis* **46**, 74–80 (2008).
- Wang, Y. *et al.* Ephrin-B2 controls VEGF-induced angiogenesis and lymphangiogenesis. *Nature* **465**, 483–486 (2010).
- Koch, U. *et al.* Delta-like 4 is the essential, nonredundant ligand for Notch1 during thymic T cell lineage commitment. *J. Exp. Med.* **205**, 2515–2523 (2008).
- Han, H. *et al.* Inducible gene knockout of transcription factor recombination signal binding protein-J reveals its essential role in T versus B lineage decision. *Int. Immunol.* **14**, 637–645 (2002).
- Harrington, L. S. *et al.* Regulation of multiple angiogenic pathways by Dll4 and Notch in human umbilical vein endothelial cells. *Microvasc. Res.* **75**, 144–154 (2008).
- Taylor, K. L., Henderson, A. M. & Hughes, C. C. Notch activation during endothelial cell network formation *in vitro* targets the basic HLH transcription factor HESR-1 and downregulates VEGFR-2/KDR expression. *Microvasc. Res.* **64**, 372–383 (2002).
- Sainson, R. C. *et al.* Cell-autonomous notch signaling regulates endothelial cell branching and proliferation during vascular tubulogenesis. *FASEB J.* **19**, 1027–1029 (2005).
- Hogan, B. M. *et al.* Vegf/Flt4 signalling is suppressed by Dll4 in developing zebrafish intersegmental arteries. *Development* **136**, 4001–4009 (2009).
- Zhang, L. *et al.* VEGFR-3 ligand-binding and kinase activity are required for lymphangiogenesis but not for angiogenesis. *Cell Res.* **20**, 1313–1331 (2010).
- Pytowski, B. *et al.* Complete and specific inhibition of adult lymphatic regeneration by a novel VEGFR-3 neutralizing antibody. *J. Natl Cancer Inst.* **97**, 14–21 (2005).
- Kirkin, V. *et al.* Characterization of indolinones which preferentially inhibit VEGF-C- and VEGF-D-induced activation of VEGFR-3 rather than VEGFR-2. *Eur. J. Biochem.* **268**, 5530–5540 (2001).
- Tammela, T. *et al.* VEGFR-3 controls tip to stalk conversion at vessel fusion sites by reinforcing Notch signalling. *Nature Cell Biol.* **13**, 1202–1213 (2011).
- Galvagni, F. *et al.* Endothelial cell adhesion to the extracellular matrix induces c-Src-dependent VEGFR-3 phosphorylation without the activation of the receptor intrinsic kinase activity. *Circ. Res.* **106**, 1839–1848 (2010).
- Stenzel, D. *et al.* Endothelial basement membrane limits tip cell formation by inducing Dll4/Notch signalling *in vivo*. *EMBO Rep.* **12**, 1135–1143 (2011).
- Estrach, S. *et al.* Laminin-binding integrins induce Dll4 expression and Notch signaling in endothelial cells. *Circ. Res.* **109**, 172–182 (2011).
- Lemmon, M. A. & Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **141**, 1117–1134 (2010).
- Lux, A., Ilacer, H., Heussen, F. M. & Joussen, A. M. Non-responders to bevacizumab (Avastin) therapy of choroidal neovascular lesions. *Br. J. Ophthalmol.* **91**, 1318–1322 (2007).
- Jubb, A. M. & Harris, A. L. Biomarkers to predict the clinical efficacy of bevacizumab in cancer. *Lancet Oncol.* **11**, 1172–1183 (2010).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank M. Schiller, M. Ehling, M. Pitulescu and M. Nakayama for the help with experiments and discussions, and G. Breier and T. Honjo for floxed *Vegfr2* and *Rbpj* mutant mice, respectively. Funding was provided by the Max Planck Society, the University of Münster and the German Research Foundation (programmes SFB 629 and SPP 1190).

**Author Contributions** R.B. and R.H.A. designed the experiments, interpreted the results and wrote the manuscript. R.B. generated and characterized the mutant mouse lines. R.B. directed M.W. and M.Z. and carried out the immunohistochemistry, immunoblots, qRT-PCR, confocal imaging and quantifications. S.F.R. developed the immunoprecipitation and immunoblotting assays and carried out the confocal imaging and quantifications. A.D. and F.R. provided the *Dll4*<sup>flxed</sup> and *Notch1*<sup>flxed</sup> mice, respectively. O.C. and B.P. generated and provided the monoclonal VEGFR2- and VEGFR3-blocking antibodies (ImClone Systems).

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details accompany the full-text HTML version of the paper at [www.nature.com/nature](http://www.nature.com/nature). Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to R.H.A. ([ralf.adams@mpi-muenster.mpg.de](mailto:ralf.adams@mpi-muenster.mpg.de)) and R.B. ([ruibenedito@mpi-muenster.mpg.de](mailto:ruibenedito@mpi-muenster.mpg.de)).

## METHODS

### Mutant mice, inducible genetic experiments and pharmacological inhibition.

To delete *Vegfr2* in the postnatal vasculature, we interbred *Vegfr2*<sup>flox</sup> mice<sup>12</sup> with transgenic mice expressing the tamoxifen-inducible recombinase CreERT2 under the control of the endothelial *Cdh5* promoter<sup>14</sup>. *Vegfr2*<sup>flox</sup>/*flox* *Cdh5*(PAC)-*CreERT2*<sup>T/+</sup> males were interbred with *Vegfr2*<sup>flox</sup>/*flox* females to generate litters containing *Vegfr2*<sup>flox</sup>/*flox* *Cdh5*(PAC)-*CreERT2*<sup>T/+</sup> (*Vegfr2*<sup>ΔEC</sup>) and control (*Vegfr2*<sup>flox</sup>/*flox*) littermates. Cre activity and gene deletion were induced by intraperitoneal injections with 100 μg tamoxifen (Sigma, T5648; 2 mg ml<sup>-1</sup>) in every pup of the litter, every 24 h at P3, P4 and P5 (Fig. 1). Notch signalling was inhibited in half of the pups by subcutaneous injection at P5 and P6 with 100 mg kg<sup>-1</sup> DAPT (Calbiochem) dissolved in 10% ethanol and 90% peanut oil. Retinas were collected at P7, 48 h after the first injection of DAPT. Control mice were injected with vehicle only. We also induced *Vegfr2* deletion from P1 to P3 and analysed tissues at P6 (Figs 2 and 3). In addition, Notch signalling was inhibited in half of the pups by subcutaneous injection of DAPT at P4 and P5 (Fig. 3d).

For combined endothelial-cell-specific loss-of-function of *Dll4* and *Vegfr2*, we interbred *Dll4*<sup>flox</sup> (ref. 15) with *Vegfr2*<sup>flox</sup> (ref. 12) and *Pdgfrb-iCre* mice<sup>13</sup>. Because *Dll4* is a haploinsufficient gene<sup>2,31</sup>, it was necessary to compare the vascular phenotypes of animals from two different types of breeding. Specifically, we interbred *Dll4*<sup>flox</sup>/*flox* *Vegfr2*<sup>flox</sup>/*flox* *Pdgfrb-iCre*<sup>T/+</sup> males with *Dll4*<sup>flox</sup>/*flox* *Vegfr2*<sup>flox</sup>/*flox* females to generate control (*Pdgfrb-iCre*<sup>T/+</sup>), *Dll4*<sup>flox</sup>/*flox* *Vegfr2*<sup>flox</sup>/*flox* *Pdgfrb-iCre*<sup>T/+</sup> (*Dll4*<sup>ΔEC</sup>), and *Dll4*<sup>flox</sup>/*flox* *Vegfr2*<sup>flox</sup>/*flox* *Pdgfrb-iCre*<sup>T/+</sup> (*Dll4*<sup>ΔEC</sup>/*Vegfr2*<sup>ΔEC</sup>) littermates. To generate *Dll4*<sup>T/+</sup> *Vegfr2*<sup>flox</sup>/*flox* *Pdgfrb-iCre*<sup>T/+</sup> (*Vegfr2*<sup>ΔEC</sup>) in Fig. 1c) and matching littermate control animals (*Pdgfrb-iCre*<sup>T/+</sup>), we interbred *Dll4*<sup>flox</sup>/*flox* *Vegfr2*<sup>flox</sup>/*flox* *Pdgfrb-iCre*<sup>T/+</sup> males with *Dll4*<sup>T/+</sup> *Vegfr2*<sup>flox</sup>/*flox* females in the same genetic background. For the combined endothelial-cell-specific loss-of-function of *Notch1* and *Vegfr2*, we interbred *Notch1*<sup>flox</sup> mice<sup>32</sup>, *Vegfr2*<sup>flox</sup> and *Cdh5*(PAC)-*CreERT2* mice to obtain the animals with the genotypes described in Supplementary Fig. 2e. To induce Cre recombinase activity and gene deletion, animals were injected with 50 μg tamoxifen (1 mg ml<sup>-1</sup>), every 24 h, at P1, P2 and P3 and dissected at P6.

To disrupt transcriptional responses downstream of activated Notch in endothelial cells, we interbred *Rbpj*<sup>flox</sup> mice<sup>16</sup> with the *Pdgfrb-iCre* transgenics. *Rbpj*<sup>flox</sup>/*flox* *Pdgfrb-iCre*<sup>T/+</sup> males were interbred with *Rbpj*<sup>flox</sup>/*flox* females to generate litters containing *Rbpj*<sup>flox</sup>/*flox* *Pdgfrb-iCre*<sup>T/+</sup> (*Rbpj*<sup>ΔEC</sup>) and control (*Rbpj*<sup>flox</sup>/*flox*) littermates. All pups were injected with 50 μg tamoxifen (1 mg ml<sup>-1</sup>), every 24 h, from P1 to P3 and dissected at P6 (Fig. 2) or P7 (Figs 1 and 3).

For the combined pharmacological inhibition of Notch and VEGFR signalling *in vivo*, we subcutaneously injected the indicated combinations of vehicle (dimethylsulphoxide (DMSO)), the Notch signalling inhibitor DAPT (Calbiochem; 100 mg kg<sup>-1</sup>), the VEGFR3 inhibitor<sup>23</sup> (MAZ51, Sigma; 10 mg kg<sup>-1</sup>) or the pan-VEGFR tyrosine kinase inhibitor (AAL-993, Enzo Life Sciences; 25 mg kg<sup>-1</sup>)<sup>33,34</sup>. Injections of the vehicle and inhibitors DAPT and MAZ51 were done twice, at P5 and 8 h before dissecting the animals at P6: that is, for a 24-h period (16 h + 8 h) in total. Injection of the vehicle and inhibitors DAPT and AAL-993 was done at P5, 24 h before collecting the retinas at P6. When *Vegfr2* deletion was combined with chemical inhibition (Fig. 4), all of the control and *Vegfr2*<sup>ΔEC</sup> litter pups were injected with 50 μg tamoxifen (1 mg ml<sup>-1</sup>) from P1 to P3 before receiving the inhibitor injections at P4 as described above.

For the combined deletion of *Rbpj* and inhibition of VEGFR signalling *in vivo*, we injected 50 μg tamoxifen (1 mg ml<sup>-1</sup>) from P1 to P3 and then injected the monoclonal antibodies anti-VEGFR2 (ImClone Systems, DC101)<sup>35</sup>, anti-VEGFR3 (ImClone Systems, mF4-31C1)<sup>22</sup>, anti-VEGF-A (Genentech, G6-31)<sup>36</sup> or control mouse or rat IgG antibodies (Jackson ImmunoResearch) at 50 mg kg day<sup>-1</sup> from P4 to P6.

**Immunohistochemistry.** All of the immunostaining was carried out with littermate tissues processed simultaneously under the same conditions. With the exception of the isolectin B4 staining of the retinas from *Dll4/Vegfr2*<sup>flox</sup>/*flox* litters (see explanation in the previous section), no comparisons have been made between animals from different litters or processed on different days.

To analyse (and quantify) the retina vascular phenotype and to preserve intact endothelial filopodia and sprouts at the angiogenic front, whole animal eyes were fixed in 4% paraformaldehyde (PFA) at 4 °C overnight. The following day, the eyes were washed in PBS before the retinas were dissected and partially cut in four quadrants to allow subsequent flat mounting. After blocking/permeabilization in 1% BSA with 0.3% Triton for some hours at room temperature or overnight at 4 °C, the retinas were washed two times in Pblec buffer (1% Triton X-100, 1 mM CaCl<sub>2</sub>, 1 mM MgCl<sub>2</sub> and 1 mM MnCl<sub>2</sub> in PBS, pH 6.8) for 20 min and then incubated for 2 h in Pblec buffer containing biotinylated isolectin B4 (Vector Labs, 1:50). Following five washes (each 20 min) in blocking solution, retinas were incubated with Alexa-Fluor-streptavidin-conjugated antibodies (Molecular Probes, 1:100)

for 2 h, washed three times further and flat-mounted on microscope glass slides with Fluoromount-G (SouthernBiotech, 0100-01). Double or triple whole-mount immunohistochemistry was performed in retinas fixed for 2 h on ice in 2% PFA or at -20 °C in methanol. After fixation, retinas were blocked for 1 h in 1% BSA with 0.05% Tween and incubated overnight or for 2 h with isolectin B4 (1:50) and the following primary antibodies: rat anti-VEGFR2 (1:200, 555307, BD Pharmingen), goat anti-VEGFR3 (1:200, AF743, R&D Systems) or goat anti-DLL4 (1:200, AF1389, R&D Systems) antibody. For detection, suitable species-specific Alexa-Fluor-coupled secondary antibodies (1:500) were used. To detect proteins in P6 mouse lungs, organs were fixed in 4% PFA. Immunohistochemistry was performed on 100-μm vibratome sections as described above for the retinas. The following additional antibodies were used: rat anti-endomucin (from D. Vestweber) and rabbit anti-GFP-Alexa Fluor 488 (1:500, A21311, Invitrogen).

**Immunoprecipitation, immunoblotting and qRT-PCR of lung lysates.** For immunoprecipitations, freshly dissected or snap frozen lungs were lysed in lysis buffer (20 mM Tris, 1 mM EDTA, 1 mM dithiothreitol (DTT), 1% NP-40, 0.1% SDS, 0.1% deoxycholate, 150 mM NaCl, 1 mM PMSF, 1 mM Na<sub>3</sub>VO<sub>4</sub>, 5 mM NaF, phosphatase inhibitor cocktail SetV (Calbiochem, 524629) and protease inhibitor cocktail (Sigma, P2714)). Lysates were centrifuged at 4 °C for 30 min at 20,000g, and aliquots were set aside for direct input blot analysis. For the immunoprecipitation, the remaining lysates were pre-cleared for 1 h at 4 °C with protein-G-sepharose or protein-A-sepharose beads. Pre-cleared lysates were incubated with primary antibody for 2 h at 4 °C, after which the beads were added and incubated for an additional 2 h at 4 °C. Immunocomplexes were washed five times with lysis buffer (without SDS and deoxycholate) and analysed by SDS-PAGE. For the detection of phosphotyrosine after VEGFR1 immunoprecipitation, membranes were blocked with BSA, and 1 mM Na<sub>3</sub>VO<sub>4</sub> was added during the blocking and washing steps. The following primary antibodies were used for immunoprecipitations: control mouse or goat IgG, mouse anti-phosphotyrosine (4G10, 05-321, Millipore) and goat anti-VEGFR1 (AF471, R&D Systems). Subsequent detection of the immunoprecipitated protein was done with the antibodies described below. As a positive control for the detection of phospho-VEGFR3 after phosphotyrosine immunoprecipitation, P5 pups received an intraperitoneal injection of 200 ng g<sup>-1</sup> body weight VEGF-C (Peprotech) or vehicle (PBS) at 30 min before tissue dissection. As a negative control to confirm the specificity of the phospho-tyrosine immunoprecipitation, a sample was treated with lambda protein phosphatase (New England Biolabs) before incubation of the pre-cleared lysate with anti-phosphotyrosine antibodies. To determine the ligand requirement for VEGFR3 phosphorylation after deletion of *Rbpj* in endothelial cells, we intraperitoneally injected 25 mg kg<sup>-1</sup> mouse VEGFR3-Fc (743-R3-100, R&D Systems) at P5 and collected the tissue 24 h later.

For the analysis of protein levels in mutant mouse pups, the inferior left lobe or the whole lung was dissected out, transferred to a reagent tube and frozen in liquid nitrogen. On the day of the immunoblot analysis, the tissue was lysed in PL buffer (20 mM Tris-HCl, pH 8.0, 1 mM EDTA, 1 mM DTT, phosphatase inhibitor cocktail SetV, 1 mM Na<sub>3</sub>VO<sub>4</sub>, protease inhibitor cocktail (1:10), 1% Triton X-100 and 150 mM NaCl) and homogenized with a cylindrical glass pestle (Potter-Elvehjem). Tissue debris was removed by centrifugation, and the supernatant was diluted in loading buffer and analysed by SDS-PAGE and immunoblotting. The following antibodies were used for immunoblotting: rabbit anti-VEGFR2 (2479, Cell Signaling Technology), rabbit anti-phospho-VEGFR2 (2478, Cell Signaling Technology), rat anti-VEGFR3 (14-5988, eBioscience), rabbit anti-VEGFR1 (sc-316, Santa Cruz Biotechnology), rabbit anti-cleaved-Notch1 (2421, Cell Signaling Technology), rat anti-CDH5 (555289, BD Biosciences), mouse anti-phosphotyrosine (4G10, 05-321, Millipore) and rabbit anti-tubulin (T5168, Sigma).

All of the immunoblot images shown represent the average result obtained from a minimum of two different lung extracts for each group, and each extract was blotted in duplicate. The blots shown in Fig. 4c were quantified with the gel analysis function in the program ImageJ.

For the analysis of mRNA relative expression levels, total RNA was isolated with the RNeasy Mini Kit (QIAGEN), and 500 ng per reaction was used to generate cDNA with the SuperScript III First-Strand Synthesis System (Invitrogen) and oligo(dT) primers. Quantitative PCR with reverse transcription (qRT-PCR) was performed by using an ABI PRISM 7900HT Sequence Detection System. TaqMan gene expression assays for murine *Gapdh*, *Dll4*, *Hey1*, *Vegfr2*, *Vegfr3*, *Pecam1*, *Cdh5*, *Pdgfrb* and *Vegfa* were used in combination with TaqMan Gene Expression Master Mix (Applied Biosystems). Gene expression was normalized to the endogenous control *Gapdh*. The relative expression differences obtained represent the average of the results obtained for two independent animals per group from two independent litters. For each animal/cDNA, two separate qRT-PCR plates were used, with duplicate reactions for each gene.

A fraction of *Vegfr2*<sup>ΔEC</sup> *Cdh5*(PAC)-*CreERT2* mutant lungs showed reduced deletion of *Vegfr2* in the lungs and were therefore excluded from the analysis. The results shown represent mutants for which the deletion of *Vegfr2* was most efficient (around 90%, Fig. 2c).

**Quantitative analysis of the retinal vasculature.** All of the images shown are representative of the vascular phenotype observed in at least eight retinas from four pups from two distinct litters per group. All quantifications were done with Volocity (Improvion) software on high-resolution confocal images representing a thin z section of the retina stained with isolectin B4. The numbers of endothelial sprouts and filopodial extensions were quantified at the retina angiogenic front in a minimum of 12 fields (sized 1,550 μm × 1,550 μm for sprouts and 387 μm × 387 μm for filopodia) and four retinas per group. The total number of sprouts or filopodia was normalized for a standard endothelial vessel length of 1,000 μm that was measured and defined according to published protocols<sup>36</sup>. The proportion of endothelial cell area coverage was calculated in a minimum of eight fields per group. Regions of quantification were selected in 1,550 μm × 1,550 μm fields, including the whole vascular plexus, for genetic deletions starting at P1, or in only the angiogenic front (around 1,550 μm × 250 μm) for 24–72 h experiments. The isolectin-B4-stained endothelial network was selected based on the strong intensity of the fluorescent signal. For each vascularized field, the ratio of the isolectin-B4-positive area to the total area was calculated and defined as the percentage of the vascularized retina area covered by endothelial cells. Vascular progression was measured by defining a straight line from the angiogenic front to the centre of the retina for each retina quadrant in low magnification stereomicroscope pictures. A minimum of 16 quadrants belonging to four retinas per group was used for quantification. The *P* values in all figures were calculated using a two-tailed Student's *t*-test.

**Image acquisition and processing.** Stained and flat-mounted retinas were analysed at high resolution with a TCS SP5 confocal microscope (Leica) or at lower resolution with an MZ16F stereomicroscope (Leica) coupled to a digital camera (Hamamatsu C4742-95). Volocity, Photoshop (Adobe) and Illustrator (Adobe) software were used for image acquisition and processing, in compliance with Nature's guide for digital images. All of the images shown in which the immunostaining levels were compared are representative of at least six different images from three different retina immunostainings per group, where the laser excitation and confocal scanner detection settings were the same between groups.

**Cell culture and *in vitro* assays.** HUVECs and immortalized mouse endothelial cells (MECs) were isolated and cultured as described previously<sup>7</sup>. Cells (5 × 10<sup>5</sup>) were plated in each well of six-well plates coated with PBS containing 5 μg ml<sup>-1</sup> fibronectin. To some wells, 1.44 μg ml<sup>-1</sup> recombinant DLL4–His (1389-D4, R&D Systems) was added to the coating solution. DMSO or an equal volume of DAPT (10 μM) dissolved in DMSO was added to the culture medium in some wells. Cells were incubated overnight and starved for 5 h before stimulation with vehicle or 50 ng ml<sup>-1</sup> VEGF-A (Peprotech) for 8 min. Cells were lysed in 200 μl 1× SDS sample buffer and subjected to immunoblotting with rabbit anti-VEGFR2 (2479, Cell Signaling Technology), rabbit anti-phospho-VEGFR2 (478, Cell Signaling Technology), rat anti-VEGFR3 (14-5988, eBioscience), rabbit anti-cleaved-Notch1 (2421, Cell Signaling Technology), rat anti-CDH5 (555289, BD Biosciences) or rabbit anti-tubulin (T5168, Sigma) antibody.

To generate human embryonic kidney 293 (HEK293) cells stably expressing mouse VEGFR3 (R3) or a tyrosine-kinase dead version (R3<sup>TK-</sup>), we first subcloned in-frame the mouse *Vegfr3* cDNA without the 3' untranslated region (UTR) into the vector pcDNA3.1 membrane-Tomato-2A. A single point mutation

(Ile1053Phe) in VEGFR3 was generated following the procedures described in the QuikChange II Site-Directed Mutagenesis Kit (Stratagene). To generate HEK293 cells stably expressing mouse VEGFR2 (R2), we first subcloned in-frame the mouse *Vegfr2* cDNA without the 3' UTR and stop codon in the vector pcDNA3.1-2A-MYFP. HEK293 cells were transfected with one of the three constructs by using Lipofectamine 2000 (Invitrogen) and selected in 500 μg ml<sup>-1</sup> G418 (neomycin) for 2 weeks. Following selection, isolated red or yellow fluorescent clones were picked and expanded. Immunoblot analysis was subsequently performed to confirm the expression of VEGFR3 or VEGFR2. For stimulation assays, 5 × 10<sup>5</sup> cells were plated in each well of six-well plates, cultured for 1 day to attach and then, on the following day, starved overnight (18 h). Starved cells were stimulated with 500 ng ml<sup>-1</sup> VEGF-C, VEGF-A (Peprotech) or vehicle, in the absence or presence of the indicated concentrations of anti-VEGFR3 (mF4-31C1) or MAZ51. Cells were lysed in 1× lysis buffer (9803, Cell Signaling Technology), and after removing the residual cell debris, the extracts were used for immunoblotting as described previously or for enzyme-linked immunosorbent assays (ELISAs). ELISA plates (96 wells) were coated overnight in capture antibody solution (PBS containing 1 μg ml<sup>-1</sup> goat-anti mouse VEGFR3 (AF743, R&D Systems) or PBS containing 1 μg ml<sup>-1</sup> goat-anti mouse VEGFR2 (AF644, R&D Systems)). The next day, the solution was removed, and blocking solution (PBS containing 0.5% BSA and 0.05% Tween 20) was added for 1 h. After three washes with wash buffer (PBS containing 0.05% Tween 20), 100 μl lysate was added in triplicate to the coated plates, which were then incubated for 2 h. The plates were washed four times in wash buffer, and primary antibodies diluted in blocking buffer were added for 2 h. For the detection of the captured phosphorylated VEGFR3 or VEGFR2 proteins, biotinylated mouse anti-phosphotyrosine (16-103, Millipore) was used. For the detection of all captured VEGFR3 and VEGFR2, we used rat anti-VEGFR3 (14-5988, eBioscience) and rat anti-VEGFR2 (2479, Cell Signaling Technology), respectively. Following four additional washing steps, species-specific or streptavidin horseradish-peroxidase-conjugated secondary antibodies were diluted in blocking solution and then incubated with the plates for 30 min. The washing procedure was repeated and TMB substrate (7004S, Cell Signaling Technology) was used to obtain the detection signal. Absorbance measurements at 450 nm and 650 nm and the relative quantification were performed on a Synergy 2 Biotek microplate reader. The ELISAs were performed in triplicate and for three times per condition. The ratio of the detected phosphorylated VEGFR3 to the total VEGFR3 (and the VEGFR2 ratio) was determined for each condition, and the relative difference from the non-treated condition was calculated. The *P* values were calculated using a two-tailed Student's *t*-test.

- Duarte, A. *et al.* Dosage-sensitive requirement for mouse Dll4 in artery development. *Genes Dev.* **18**, 2474–2478 (2004).
- Radtke, F. *et al.* Deficient T cell fate specification in mice with an induced inactivation of Notch1. *Immunity* **10**, 547–558 (1999).
- Manley, P. W. *et al.* Advances in the structural biology, design and clinical development of VEGF-R kinase inhibitors for the treatment of angiogenesis. *Biochim. Biophys. Acta* **1697**, 17–27 (2004).
- Manley, P. W. *et al.* Anthranilic acid amides: a novel class of antiangiogenic VEGF receptor kinase inhibitors. *J. Med. Chem.* **45**, 5687–5693 (2002).
- Prewett, M. *et al.* Antivascular endothelial growth factor receptor (fetal liver kinase 1) monoclonal antibody inhibits tumor angiogenesis and growth of several mouse and human tumors. *Cancer Res.* **59**, 5209–5218 (1999).
- Liang, W. C. *et al.* Cross-species vascular endothelial growth factor (VEGF)-blocking antibodies completely inhibit the growth of human tumor xenografts and measure the contribution of stromal VEGF. *J. Biol. Chem.* **281**, 951–961 (2006).



# The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier–Gorlin syndrome

Alex J. Kuo<sup>1\*</sup>, Jikui Song<sup>2†\*</sup>, Peggie Cheung<sup>1\*</sup>, Satoko Ishibe–Murakami<sup>2</sup>, Sayumi Yamazoe<sup>3</sup>, James K. Chen<sup>3</sup>, Dinshaw J. Patel<sup>2</sup> & Or Gozani<sup>1</sup>

The recognition of distinctly modified histones by specialized ‘effector’ proteins constitutes a key mechanism for transducing molecular events at chromatin to biological outcomes<sup>1</sup>. Effector proteins influence DNA-templated processes, including transcription, DNA recombination and DNA repair; however, no effector functions have yet been identified within the mammalian machinery that regulate DNA replication. Here we show that ORC1—a component of ORC (origin of replication complex), which mediates pre-DNA replication licensing<sup>2</sup>—contains a bromo adjacent homology (BAH) domain that specifically recognizes histone H4 dimethylated at lysine 20 (H4K20me2). Recognition of H4K20me2 is a property common to BAH domains present within diverse metazoan ORC1 proteins. Structural studies reveal that the specificity of the BAH domain for H4K20me2 is mediated by a dynamic aromatic dimethyl-lysine-binding cage and multiple intermolecular contacts involving the bound peptide. H4K20me2 is enriched at replication origins, and abrogating ORC1 recognition of H4K20me2 in cells impairs ORC1 occupancy at replication origins, ORC chromatin loading and cell-cycle progression. Mutation of the ORC1 BAH domain has been implicated in the aetiology of Meier–Gorlin syndrome (MGS)<sup>3,4</sup>, a form of primordial dwarfism<sup>5</sup>, and ORC1 depletion in zebrafish results in an MGS-like phenotype<sup>4</sup>. We find that wild-type human ORC1, but not ORC1–H4K20me2-binding mutants, rescues the growth retardation of *orc1* morphants. Moreover, zebrafish depleted of H4K20me2 have diminished body size, mirroring the phenotype of *orc1* morphants. Together, our results identify the BAH domain as a novel methyl-lysine-binding module, thereby establishing the first direct link between histone methylation and the metazoan DNA replication machinery, and defining a pivotal aetiological role for the canonical H4K20me2 mark, via ORC1, in primordial dwarfism.

The identification of protein modules that recognize the broad spectrum of modifications present on histone proteins is critical for understanding how chromatin dynamics influence fundamental nuclear processes<sup>6</sup>. The BAH domain is an evolutionarily conserved chromatin-associated motif<sup>7</sup>. In budding yeast, the BAH domain of the silencing regulator Sir3 was shown to be a nucleosome-binding domain<sup>8–10</sup>, suggesting that other BAH domains might function as chromatin-effector modules. To test this hypothesis, we screened several BAH domains from yeast and human for binding activity on peptide microarrays containing 82 modified and unmodified histone peptides, including 54 uniquely methylated peptides (Supplementary Fig. 1)<sup>11</sup>. The BAH domain of human ORC1 (ref. 12) (hORC1<sub>BAH</sub>: amino acids 1–185) bound with high specificity to H4K20me2 peptides (Fig. 1a). In peptide pull-down assays, hORC1<sub>BAH</sub> bound H4K20me2 peptides, but did not interact with several other dimethylated histone peptides (Fig. 1b). In addition, hORC1<sub>BAH</sub> did not interact with unmodified H4 peptide, and showed a strong preference for H4K20me2 over H4K20me1 or H4K20me3 (Fig. 1c). Quantification of

the interaction by isothermal titration calorimetry (ITC) demonstrated that hORC1<sub>BAH</sub> bound to H4(14–25)K20me2 peptides with a dissociation constant ( $K_d$ ) of 5.2  $\mu$ M, whereas monomethylation and trimethylation at K20 reduced the binding affinity fivefold ( $K_d$  = 24.8  $\mu$ M) and eightfold ( $K_d$  = 40.0  $\mu$ M), respectively (Fig. 1d). Finally, hORC1<sub>BAH</sub> bound to full-length H4K20me2 that is present in bulk-purified histones (Fig. 1e). Together, these *in vitro* results demonstrate that hORC1<sub>BAH</sub> binds with high specificity and affinity to H4K20me2.

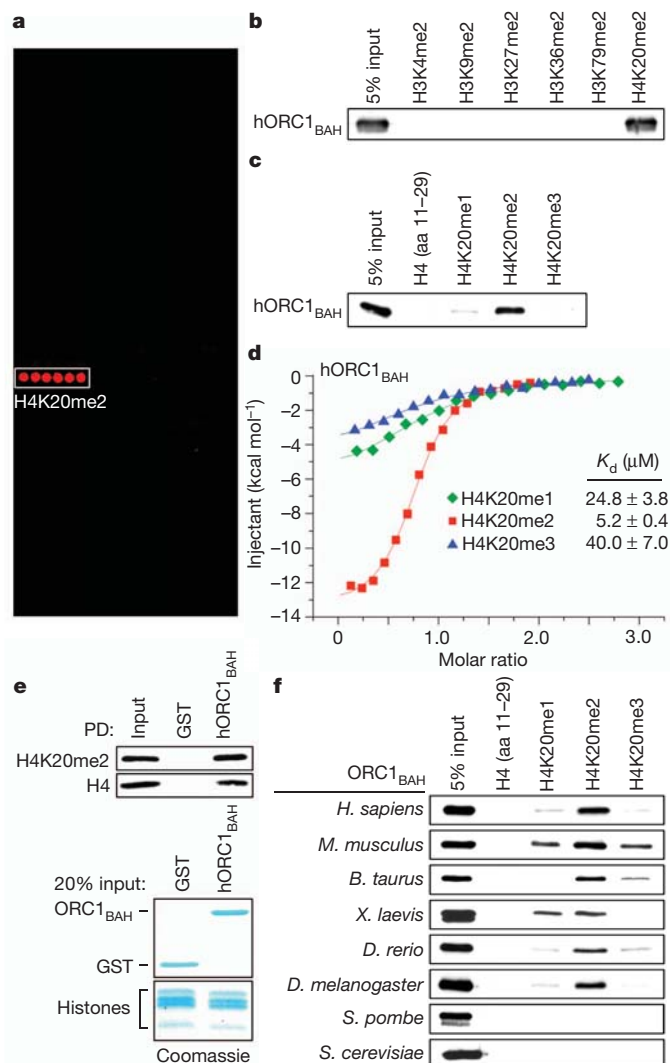
ORC1 is an evolutionarily conserved protein, found in virtually all eukaryotes<sup>2,12</sup>. As shown in Fig. 1f, H4K20me2 recognition is a common activity of BAH domains of ORC1 from several different metazoans. In contrast, the BAH domains of ORC1 from yeast *Saccharomyces cerevisiae* (scORC1<sub>BAH</sub>) and *Schizosaccharomyces pombe* lack H4K20me2-binding activity. In both of these yeast species, it is known that ORC binds directly to DNA sequences at origins of replication, whereas in metazoans, chromatin has a more significant role in directing ORC to replication origins<sup>2,13,14</sup>.

To understand the molecular basis of the ORC1 BAH domain recognition of H4K20me2, the crystal structure of the complex between the H4(14–25)K20me2 peptide and the BAH domain of mouse ORC1(9–170) (mORC1<sub>BAH</sub>) was solved at 1.95 Å resolution (Fig. 2a; crystallization statistics in Supplementary Table 1). mORC1<sub>BAH</sub> adopts a characteristic BAH-domain fold (Fig. 2a), as first reported for the scORC1<sub>BAH</sub><sup>15</sup>. Complex formation between H4(14–25)K20me2 and mORC1<sub>BAH</sub> is mediated by van der Waals contacts, hydrogen bonding and electrostatic interactions (Fig. 2b, c). The dimethylammonium group of H4K20me2 is anchored by cation– $\pi$  interactions with the side chains of an aromatic cage located at the mORC1<sub>BAH</sub> surface and composed of Tyr 63, Trp 87, Tyr 114 and Trp 119, and the dimethylammonium proton is hydrogen bonded to the side chain of Glu 93 (Fig. 2d), similar to H4K20me2 bound to the 53BP1 Tudor domain<sup>16</sup> (Supplementary Fig. 2). The aromatic cage in mORC1<sub>BAH</sub> is absent in the published structure of scORC1<sub>BAH</sub> (Supplementary Fig. 3)<sup>15</sup>, explaining why scORC1<sub>BAH</sub> does not bind H4K20me2 (Fig. 1f).

The preference of hORC1<sub>BAH</sub> for H4K20me2 over its me1/3 counterparts (see Fig. 1c) probably reflects the contribution from the hydrogen bond formed between the dimethylammonium proton and the side-chain carboxylate of Glu 93 (Fig. 2c). Indeed, previous studies demonstrated that a single Tyr-to-Glu substitution within an otherwise aromatic cage of the bromodomain and PHD domain transcription factor (BPTF) plant homeo domain (PHD) finger reverses the binding preference from trimethyl- to dimethyl-lysine<sup>17</sup>. Engineering a binding preference for H4K20me1 would probably require incorporation of a second carboxylate group to line the aromatic cage of hORC1<sub>BAH</sub> and facilitate hydrogen-bond formation with both monomethylammonium protons. The high degree of specificity of mORC1<sub>BAH</sub> for the sequence surrounding H4K20 is conferred by a series of intermolecular

<sup>1</sup>Department of Biology, Stanford University, Stanford, California 94305, USA. <sup>2</sup>Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. <sup>3</sup>Department of Chemical and Systems Biology, Stanford School of Medicine, Stanford, California 94305, USA. <sup>†</sup>Present address: Department of Biochemistry, University of California, Riverside, California 92521, USA.

\*These authors contributed equally to this work.



**Figure 1 | The ORC1 BAH domain is a novel H4K20me2-binding module.**

**a**, hORC1<sub>BAH</sub> preferentially binds H4K20me2 peptides. Microarrays spotted with 82 distinct histone peptides as indicated in Supplementary Fig. 1 were probed with glutathione S-transferase (GST) fused to hORC1<sub>BAH</sub>. Red spots indicate positive binding. **b**, **c**, hORC1<sub>BAH</sub> specifically binds H4K20me2 peptides. Western blot analysis of histone peptide pull-downs with GST-hORC1<sub>BAH</sub> and the indicated biotinylated peptides. aa, amino acids. **d**, hORC1<sub>BAH</sub> binds with highest affinity to H4K20me2. ITC was used to determine the  $K_d$  values for the interaction of hORC1<sub>BAH</sub> with the indicated peptides: H4K20me1 (green diamonds), H4K20me2 (red squares) and H4K20me3 (blue triangles). Standard deviation (s.d.) was derived from nonlinear fitting. **e**, hORC1<sub>BAH</sub> binds full-length H4K20me2. Top, western blot of GST-hORC1<sub>BAH</sub> and GST control pull-downs from calf thymus histones (CTH) with the indicated antibodies. Bottom, Coomassie blue stain of input (20% of total). PD, GST pull-downs. **f**, H4K20me2 binding is a common property of ORC1 BAH domains from diverse metazoan species. Binding assays as in **b** with GST-fused ORC1 BAH domains from the indicated species and using the indicated peptides.

hydrogen-bonding interactions involving Lys 16, His 18, Val 21 and Arg 23 of the bound peptide and residues lining the peptide-binding channel of mORC1<sub>BAH</sub> (Fig. 2b, c). Although backbone-backbone interactions represent the majority of contacts, a water-mediated hydrogen bond involving the imidazole ring of His 18, as well as a salt bridge between Lys 16 of the H4 peptide and Glu 126 of mORC1<sub>BAH</sub>, are also observed (Fig. 2b, c). In this regard, acetylation at K16 slightly reduced the binding affinity of hORC1<sub>BAH</sub> and mORC1<sub>BAH</sub> for the H4K16acK20me2 peptide (Supplementary Fig. 4).

To investigate the structural dynamics of H4K20me2 binding, we determined the 1.70 Å crystal structure of mORC1<sub>BAH</sub> in the free state

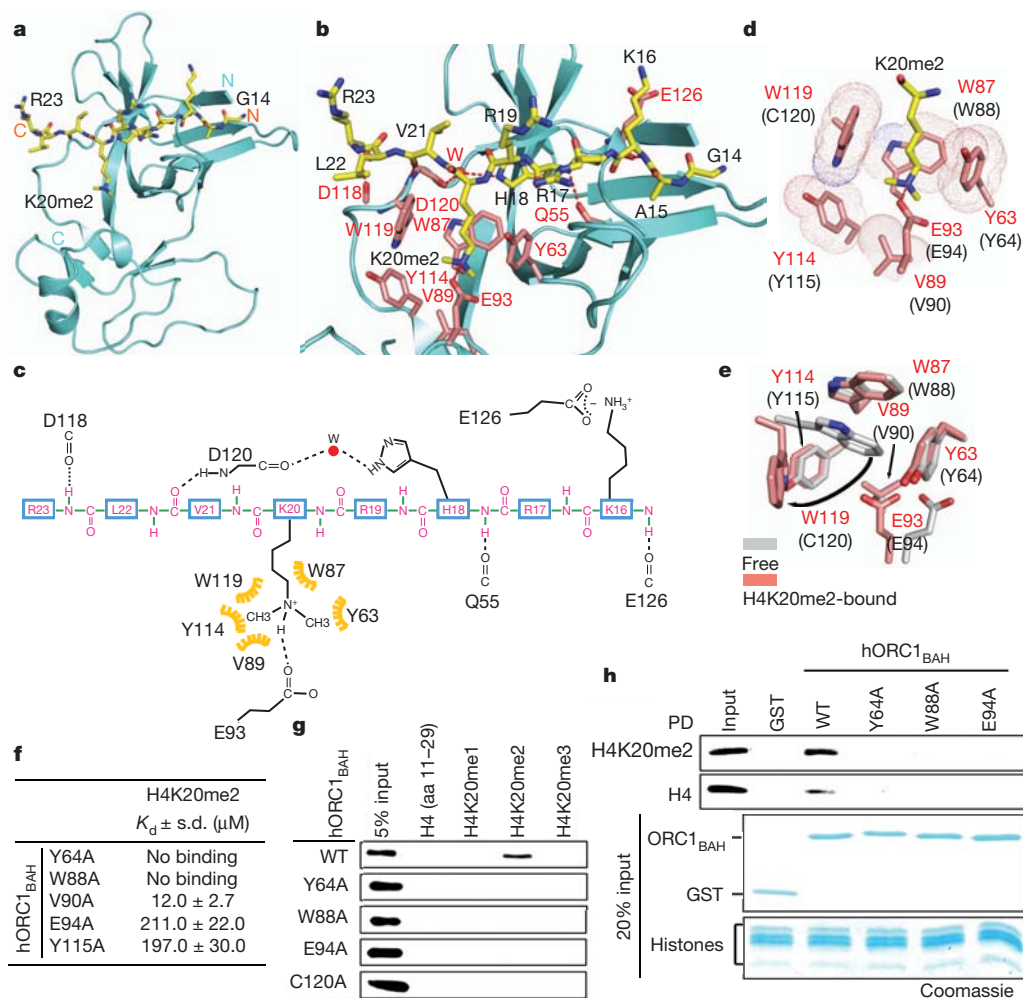
(crystallization statistics in Supplementary Table 1). Superimposition of the mORC1<sub>BAH</sub> structures in the free and peptide-bound states demonstrated that, although the interaction with H4K20me2 does not alter the conformation for the majority of the protein, the K20me2-binding pocket of mORC1<sub>BAH</sub> undergoes a significant rearrangement upon complex formation (Fig. 2e and Supplementary Fig. 5). Specifically, formation of the mORC1<sub>BAH</sub>-H4K20me2 complex prompts the indole ring of Trp 119, which in the free state stacks with the indole ring of Trp 87, to swing aside and accommodate insertion of H4K20me2 (Fig. 2e). In addition, the interaction elicits reorientation of Glu 93 to allow its side chain to hydrogen bond with the methylammonium proton of H4K20me2 (Fig. 2e).

ITC studies establish that mORC1<sub>BAH</sub> binds to H4K20me2 peptide with a  $K_d$  of 9.6  $\mu$ M, with the binding affinity towards H4K20me1 and H4K20me3 weaker by a factor of 2 to 3 ( $K_d$  = 32.7  $\mu$ M and 17.4  $\mu$ M, respectively; Supplementary Fig. 6a). The mORC1<sub>BAH</sub> domain shows moderate sequence conservation through evolution, with relatively high conservation of residues lining the H4K20me2-binding pocket (Supplementary Fig. 6b); hORC1<sub>BAH</sub> retains all the aromatic cage residues that recognize H4K20me2, except for Trp 119 of mORC1<sub>BAH</sub>, which is substituted by Cys 120 in hORC1<sub>BAH</sub>. Replacing mORC1<sub>BAH</sub> Trp 119 with a cysteine (human counterpart) resulted in increased discrimination of H4K20me2 over H4K20me3 from twofold to threefold, approaching the eightfold discrimination observed for hORC1<sub>BAH</sub> (Supplementary Fig. 6a). However, given that this mutant could not fully recapitulate the H4K20me2-binding specificity of hORC1<sub>BAH</sub> (Supplementary Fig. 6a), other residues probably contribute to H4K20me2 discrimination. Indeed, a C120A substitution in hORC1<sub>BAH</sub> decreases binding to all three methylation states of H4K20me, with retention of H4K20me2 specificity (Supplementary Fig. 6a).

ITC analysis also demonstrated that alanine substitutions of the remaining residues lining the H4K20me2-binding pocket of hORC1<sub>BAH</sub> (Y64A, W88A, E94A and Y115A) abolished or largely diminished H4K20me2 recognition (Fig. 2f). Similar results were observed with hORC1<sub>BAH</sub> and mORC1<sub>BAH</sub> cage mutants in peptide-binding assays (Fig. 2g and Supplementary Fig. 6c) and in histone-binding assays (Fig. 2h). These results further support the structural analysis and establish the molecular basis for the H4K20me2-ORC1<sub>BAH</sub> interaction.

H4K20me2 is an abundant H4 modification<sup>18</sup>, and the BAH domain of ORC1, although dispensable for ORC complex assembly, has been shown to be important for loading of the complex onto chromatin in human cells<sup>19</sup>. We therefore postulated that the interaction between ORC1 and H4K20me2 might regulate ORC stabilization at chromatin. First, ORC components (ORC2, ORC3 and ORC5) affinity purified with two structure-guided H4K20me2-binding mutants (hORC1(Y64A) and hORC1(W88A)) with an efficiency equal to that observed with the wild-type protein in both U2OS (Fig. 3a) and HT1080 (Supplementary Fig. 7a) human cell lines, suggesting that H4K20me2 binding by ORC1 is dispensable for ORC complex assembly. Next, analysis of lysates biochemically separated into chromatin-enriched and soluble fractions from cells stably expressing hORC1, hORC1(Y64A) or hORC1(W88A) demonstrated that hORC1(Y64A) and hORC1(W88A) enrichment at chromatin was considerably reduced in comparison to hORC1 (Fig. 3b and Supplementary Fig. 7b). Moreover, chromatin association of the ORC components ORC2, ORC3, ORC5 and ORC6 was compromised in cells expressing the ORC1 H4K20me2-binding-pocket mutants (Fig. 3c and Supplementary Fig. 7c). Thus, hORC1<sub>BAH</sub> binding to H4K20me2 is required for efficient stabilization of ORC1 and other ORC components at chromatin.

Local chromatin structure is thought to have a role in the mechanism that determines metazoan origins of replication<sup>13</sup>. In this context, chromatin immunoprecipitation (ChIP) assays with a highly specific H4K20me2 antibody in G1-synchronized U2OS cells (Supplementary Figs 8 and 9) demonstrated an increase in the H4K20me2 signal at two defined human replication origins<sup>20–23</sup> relative to adjacent



**Figure 2** | The molecular basis of H4K20me2 recognition by ORC1<sub>BAH</sub>. **a–c**, 1.95 Å crystal structure of mORC1<sub>BAH</sub> complexed with H4(14–25)K20me2 peptide. **a**, Ribbon representation of mORC1<sub>BAH</sub> bound to H4K20me2 peptide. The mORC1<sub>BAH</sub> (cyan) and the bound H4K20me2 peptide (yellow) are shown in ribbon and stick representations, respectively. **b**, Details of intermolecular contacts in the mORC1<sub>BAH</sub>–H4K20me2 complex. mORC1<sub>BAH</sub> and H4K20me2 peptide residues are coloured in pink and yellow, respectively, with hydrogen bonds depicted as red dashed lines, and a water molecule (W) as a red sphere. **c**, Schematic representation of intermolecular contacts in the mORC1<sub>BAH</sub>–H4K20me2 complex. The residues from the H4K20me2 peptide and mORC1<sub>BAH</sub> are coloured in magenta and black, respectively. Yellow indicates the hydrophobic contact. **d**, Positioning of the K20me2 side chain within an aromatic cage of the indicated residues (red) on

the surface of mORC1<sub>BAH</sub>. The equivalent cage residues in hORC1<sub>BAH</sub> are labelled in parentheses. **e**, Structural overlay of the mORC1<sub>BAH</sub> K20me2-binding aromatic cage in the free (silver) and H4K20me2-bound states (salmon). The curved arrow indicates binding-induced structural shift. **f**, Mutations in the hORC1<sub>BAH</sub> H4K20me2-binding channel impair H4K20me2 recognition. ITC analysis of the indicated hORC1<sub>BAH</sub> mutants binding to H4K20me2 peptide; s.d. derived from nonlinear fitting. **g**, **h**, Mutations in the hORC1<sub>BAH</sub> dimethyl-lysine-binding cage abrogate H4K20me2 recognition. **g**, Binding assays as in Fig. 1b with the indicated hORC1<sub>BAH</sub> mutant proteins and biotinylated peptides. **h**, Top, western blot analysis of CTH-binding assays as in Fig. 1e with the indicated proteins and antibodies. PD, GST pull-downs. Bottom, Coomassie blue stain of input GST-fusion proteins and histones (20% of total). WT, wild type.

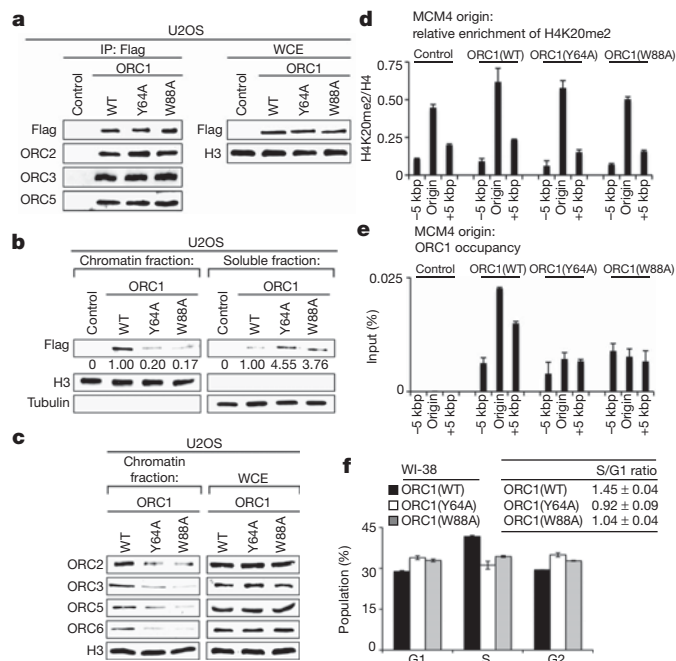
sequences (Fig. 3d and Supplementary Fig. 10). Moreover, like the pattern of H4K20me2, hORC1 occupancy peaked at origins relative to adjacent sequences (Fig. 3e and Supplementary Fig. 10). In contrast, hORC1(Y64A) and hORC1(W88A) enrichment at origins was not observed, even though the H4K20me2 peak at origins is present in these cell lines (Figs 3d, e and Supplementary Fig. 10). These data suggest that H4K20me2 may have a role at human origins by stabilizing ORC1 at chromatin.

ORC marks genomic origins of replication during the G1 phase of the cell cycle and serves to nucleate assembly of the pre-DNA replication complex (pre-RC) to coordinate DNA replication during S phase<sup>14</sup>. As a core component of this licensing machinery, ORC1 is required for efficient cell-cycle progression and the transition from G1 to S phase<sup>4,19</sup>. In this context, cell-cycle analysis demonstrated a decreased S-phase population and a lower S/G1 ratio in WI-38 primary fibroblasts transiently expressing the H4K20me2-binding mutants hORC1(Y64A) and hORC1(W88A) relative to wild-type

hORC1-overexpressing cells (Fig. 3f). Notably, mouse embryonic fibroblasts isolated from mice lacking the enzymes that generate H4K20me2 (Suv420h1 and Suv420h2) show DNA replication defects, including delayed S-phase entry<sup>18</sup>. Thus, ORC1 recognition of the canonical H4K20me2 mark appears to facilitate ORC chromatin loading to promote DNA replication and cell-cycle progression.

Mutations in components of the pre-DNA replication machinery, including ORC1, have recently been identified in individuals with MGS and, in a zebrafish model, *orc1* morphants display an MGS-like proportionate dwarfism phenotype<sup>3,4,24</sup>. In MGS individuals, the majority of ORC1 mutations fall within the protein's BAH domain<sup>3,4</sup>, including two mutants (F89S and E127G) that are in regions that impinge upon H4K20me2 recognition (Fig. 4a) and decrease binding of hORC1<sub>BAH</sub> to H4K20me2 peptides (Fig. 4b, c). Furthermore, mice deficient in H4K20me2 synthesis are born significantly smaller than control littermates<sup>18</sup>. Together, these observations implicate disruption of the ORC1–H4K20me2 interaction in MGS pathogenesis. To

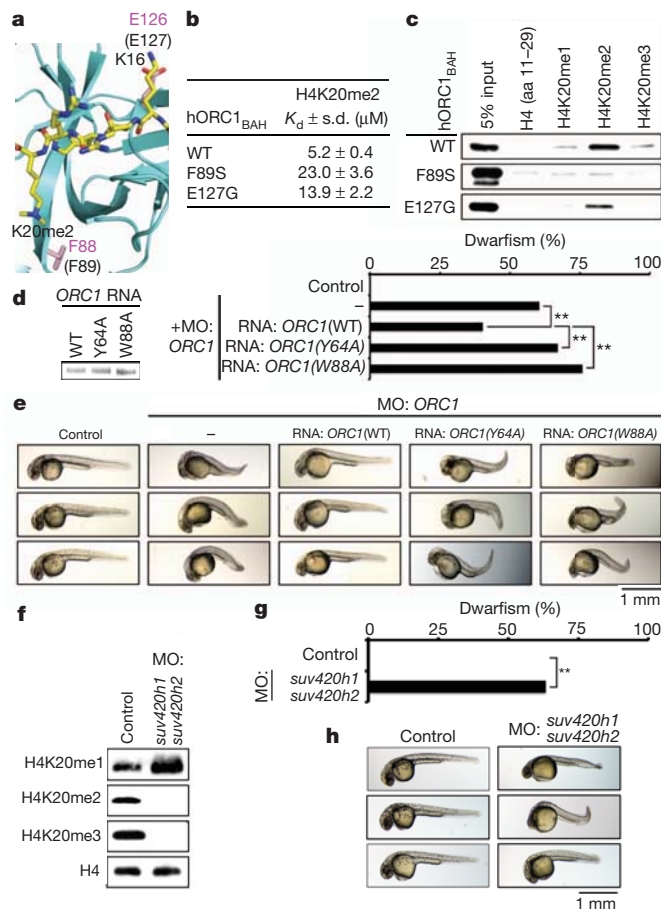




**Figure 3 | ORC1–H4K20me2 interaction regulates ORC chromatin association and cell-cycle progression.** **a**, Western blot analysis with the indicated antibodies of wild-type (WT) and H4K20me2-binding-pocket mutants (Y64A and W88A) affinity-purified Flag-tagged ORC1 complexes from U2OS cells. Control, empty vector control immunoprecipitate (IP). WCE, whole cell extract. **b**, The ORC1<sub>BAH</sub>–H4K20me2 interaction is required for efficient ORC1 chromatin association. Western blot analysis of lysates biochemically separated into chromatin-enriched and soluble fractions from U2OS cells stably expressing the indicated ORC1 protein. Quantification of Flag–ORC1 levels is shown. Control, empty vector control lysates. Tubulin and H3 levels are shown as control for the integrity of fractionation. **c**, Disruption of ORC1 binding to H4K20me2 destabilizes ORC chromatin association. Western blot analysis of biochemically purified chromatin from U2OS cells as in **b** with the indicated antibodies. Total ORC protein levels in WCE are shown. **d**, H4K20me2 is enriched at DNA replication origins. H4K20me2 signal normalized to total H4 at the MCM4 origin and indicated flanking regions in G1-phase-synchronized U2OS cells stably expressing the indicated ORC protein; y-axis shows H4K20me2 ChIP/H4 ChIP. **e**, An intact BAH domain is required for ORC1 occupancy at replication origins. Occupancy of Flag-tagged hORC1, hORC1(Y64A), hORC1(W88A) or control was determined by ChIP analysis (y-axis shows percentage input) as in **d**. Error bars in **d** and **e** indicate s.e.m. from three experiments. **f**, The ORC1–H4K20me2 interaction is required for efficient cell-cycle progression. The cell-cycle profile of WI-38 cells transiently expressing GFP and the indicated ORC1 protein was determined by flow cytometry. GFP-positive cells were used to ensure that only transfected cells were analysed. Left, percentage of cells in the indicated cell-cycle phase is shown. Right, the S/G1 ratio for the indicated transfections is shown. Error bars indicate s.d. from two experiments.

test this hypothesis, an *orc1* zebrafish morphant reconstitution system was established. Injection of two independent *orc1*-targeting morpholino oligonucleotides (MOs) resulted in growth retardation, recapitulating published results<sup>4</sup> (Fig. 4d, e and data not shown). Co-injection of human *ORC1* messenger RNA with zebrafish *orc1*-targeting MOs attenuated the dwarf phenotype observed in the *orc1* morphant alone (Fig. 4d, e). In contrast, co-injection of hORC1(Y64A) and hORC1(W88A) mRNA failed to rescue *orc1* morphants, moderately aggravating the growth retardation phenotype (Fig. 4d, e).

The H4K20me2/3 lysine methyltransferases (KMTs) Suv420h1 and Suv420h2 are conserved in zebrafish<sup>25</sup>. As shown in Fig. 4f, injection of MOs targeting *Danio rerio* *suv420h1* and *suv420h2* resulted in global depletion of H4K20me2 and H4K20me3, and an increase in H4K20me1; these changes in H4K20 methylation are similar to those observed in cells derived from *Suv420h1/20h2* double-knockout mice<sup>18</sup>. Analysis of body size in the *suv420h1/20h2* morphants demonstrated



**Figure 4 | Disruption of the ORC1<sub>BAH</sub>–H4K20me2 interaction leads to dwarfism in zebrafish.** **a–c**, MGS-associated mutations F89S and E127G impair H4K20me2 binding by hORC1. **a**, Close-up view of mORC1<sub>BAH</sub> bound to H4K20me2, with residues F88 and E126 (equivalent to F89 and E127 in hORC1, respectively) shown in stick representation. **b**, ITC analysis as in Fig. 1d of F89S and E127G hORC1<sub>BAH</sub> mutants binding to H4K20me2 peptides. ITC data for wild-type hORC1<sub>BAH</sub> from Fig. 1d. **c**, Binding assays as in Fig. 1b with the indicated hORC1<sub>BAH</sub> mutant proteins and biotinylated peptides. **d**, **e**, The H4K20me2-recognition activity of hORC1 is required to rescue the dwarfism phenotype of *orc1* morphants. **d**, Quantification of dwarf phenotype in zebrafish injected with MOs targeting the *orc1* translation start site alone or MO co-injected with the indicated hORC1 mRNAs. Insert on left shows electrophoresis analysis of the indicated hORC1 mRNAs used for reconstitution. Control, uninjected embryos. Dwarfism was defined as a reduction of body length of  $\geq 3$  s.d. relative to the average size of the control zebrafish. Zebrafish analysed: control, 26; *orc1* MO, 86; MO + hORC1 wild type, 67; MO + hORC1(Y64A), 85; MO + hORC1(W88A), 83. P values are calculated with a two-tailed unpaired Student's *t*-test. \*\**P* < 0.01. **e**, Representative images of zebrafish in **d** 1 day post-fertilization (dpf). **f**, Depletion of H4K20me2/3 in *suv420h1/suv420h2* morphants. Western blot analysis using the indicated antibodies of whole animal extracts 1 dpf from control zebrafish or zebrafish injected with MOs targeting the translation start sites of *suv420h1* and *suv420h2*. Control, uninjected embryos. **g**, **h**, Dwarfism in *suv420h1/suv420h2* morphants. **g**, Quantification of dwarfism in *suv420h1/suv420h2* morphants relative to controls as described in **d**. Zebrafish analysed: control, 38; MO, 60. \*\**P* < 0.01. **h**, Representative images of zebrafish in **g** 1 dpf.

that H4K20me2 depletion results in a significant dwarfism phenotype compared to controls (Fig. 4g, h). Finally, *orc1/suv420h1/20h2* morphants did not display significantly more dwarfisms than *orc1* or *suv420h1/20h2* morphants (Supplementary Fig. 11). Together, these data indicate that the specific recognition of H4K20me2 by the BAH domain of ORC1 has a crucial role in determining organism body size.

We have identified the BAH domain as a novel methyl-lysine-binding module, demonstrating that H4K20me2 recognition is an evolutionarily conserved function for BAH domains from diverse

metazoan ORC1 proteins. To our knowledge, these results provide the first direct link between histone methylation and the metazoan DNA replication machinery. The mechanism that determines metazoan origins of replication is thought to be dependent upon both the information encoded in the DNA sequence at origins, as well as the local chromatin architecture<sup>13,14</sup>. Several histone modifications have been detected at origins<sup>20–22,26,27</sup>, including H3K27me1, which regulates replication of DNA at heterochromatin in *Arabidopsis*<sup>28</sup>. The mark H4K20me1 is present in early G1 at human origins and probably serves as the chromatin template for H4K20me2 catalysis<sup>18,29,30</sup>. Because H4K20me1 is generated in a highly cell-cycle-regulated manner by the KMT PR-Set7 (also known as SET8), this modification may regulate licensing by governing the temporal and spatial availability of H4K20me2 at origins<sup>29</sup>. We postulate that the recognition of H4K20me2 by ORC1 cooperates with other ORC chromatin-loading mechanisms in marking replication origins. On the basis of this model, during developmental phases requiring rapid cell division, inefficient pre-RC formation due to abrogation of the ORC1–H4K20me2 interaction would result in delayed cell-cycle progression and insufficient cellular proliferation, a characteristic of proportional ‘hypocellular’ dwarfism disorders like MGS<sup>5</sup>. Together, our findings reveal a new function for histone methylation signalling at chromatin in the regulation of DNA replication and organismal growth.

## METHODS SUMMARY

Materials, peptide microarray experiments, binding assays, structure analysis and crystallization conditions, and zebrafish experiments and data analysis are described in detail in Methods. For the zebrafish experiments, wild-type AB zebrafish (*D. rerio*) were maintained and raised using standard protocols. All zebrafish were treated in accordance with AAALAC approved guidelines at Stanford University (protocol number 10511).

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 24 November 2011; accepted 17 February 2012.**

**Published online 7 March 2012.**

1. Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D. & Patel, D. J. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nature Struct. Mol. Biol.* **14**, 1025–1040 (2007).
2. Duncker, B. P., Chesnokov, I. N. & McConkey, B. J. The origin recognition complex protein family. *Genome Biol.* **10**, 214 (2009).
3. Bicknell, L. S. *et al.* Mutations in the pre-replication complex cause Meier-Gorlin syndrome. *Nature Genet.* **43**, 356–359 (2011).
4. Bicknell, L. S. *et al.* Mutations in ORC1, encoding the largest subunit of the origin recognition complex, cause microcephalic primordial dwarfism resembling Meier-Gorlin syndrome. *Nature Genet.* **43**, 350–355 (2011).
5. Klingseisen, A. & Jackson, A. P. Mechanisms and pathways of growth failure in primordial dwarfism. *Genes Dev.* **25**, 2011–2024 (2011).
6. Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693–705 (2007).
7. Callebaut, I., Courvalin, J. C. & Moron, J. P. The BAH (bromo-adjacent homology) domain: a link between DNA methylation, replication and transcriptional regulation. *FEBS Lett.* **446**, 189–193 (1999).
8. Onishi, M., Liou, G. G., Buchberger, J. R., Walz, T. & Moazed, D. Role of the conserved Sir3-BAH domain in nucleosome binding and silent chromatin assembly. *Mol. Cell* **28**, 1015–1028 (2007).
9. Sampath, V. *et al.* Mutational analysis of the Sir3 BAH domain reveals multiple points of interaction with nucleosomes. *Mol. Cell Biol.* **29**, 2532–2545 (2009).
10. Armache, K. J., Garlick, J. D., Canzio, D., Narlikar, G. J. & Kingston, R. E. Structural basis of silencing: Sir3 BAH domain in complex with a nucleosome at 3.0 Å resolution. *Science* **334**, 977–982 (2011).

11. Bua, D. J. *et al.* Epigenome microarray platform for proteome-wide dissection of chromatin-signaling networks. *PLoS ONE* **4**, e6789 (2009).
12. Bell, S. P., Mitchell, J., Leber, J., Kobayashi, R. & Stillman, B. The multidomain structure of Orc1p reveals similarity to regulators of DNA replication and transcriptional silencing. *Cell* **83**, 563–568 (1995).
13. Dorn, E. S. & Cook, J. G. Nucleosomes in the neighborhood: new roles for chromatin modifications in replication origin control. *Epigenetics* **6**, 552–559 (2011).
14. Bell, S. P. & Dutta, A. DNA replication in eukaryotic cells. *Annu. Rev. Biochem.* **71**, 333–374 (2002).
15. Zhang, Z., Hayashi, M. K., Merkel, O., Stillman, B. & Xu, R. M. Structure and function of the BAH-containing domain of Orc1p in epigenetic silencing. *EMBO J.* **21**, 4600–4611 (2002).
16. Botuyan, M. V. *et al.* Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. *Cell* **127**, 1361–1373 (2006).
17. Li, H. *et al.* Structural basis for lower lysine methylation state-specific readout by MBT repeats of L3MBTL1 and an engineered PHD finger. *Mol. Cell* **28**, 677–691 (2007).
18. Schotta, G. *et al.* A chromatin-wide transition to H4K20 monomethylation impairs genome integrity and programmed DNA rearrangements in the mouse. *Genes Dev.* **22**, 2048–2061 (2008).
19. Noguchi, K., Vassilev, A., Ghosh, S., Yates, J. L. & DePamphilis, M. L. The BAH domain facilitates the ability of human Orc1 protein to activate replication origins *in vivo*. *EMBO J.* **25**, 5372–5382 (2006).
20. Tardat, M. *et al.* The histone H4 Lys 20 methyltransferase PR-Set7 regulates replication origins in mammalian cells. *Nature Cell Biol.* **12**, 1086–1093 (2010).
21. Tardat, M., Murr, R., Herceg, Z., Sardet, C. & Julien, E. PR-Set7-dependent lysine methylation ensures genome replication and stability through S phase. *J. Cell Biol.* **179**, 1413–1426 (2007).
22. Miotto, B. & Struhl, K. HBO1 histone acetylase is a coactivator of the replication licensing factor Cdt1. *Genes Dev.* **22**, 2633–2638 (2008).
23. Kitsberg, D., Selig, S., Keshet, I. & Cedar, H. Replication structure of the human  $\beta$ -globin gene domain. *Nature* **366**, 588–590 (1993).
24. Guernsey, D. L. *et al.* Mutations in origin recognition complex gene *ORC4* cause Meier-Gorlin syndrome. *Nature Genet.* **43**, 360–364 (2011).
25. Sun, X. J. *et al.* Genome-wide survey and developmental expression mapping of zebrafish SET domain-containing genes. *PLoS ONE* **3**, e1499 (2008).
26. Costas, C. *et al.* Genome-wide mapping of *Arabidopsis thaliana* origins of DNA replication and their associated epigenetic marks. *Nature Struct. Mol. Biol.* **18**, 395–400 (2011).
27. Miotto, B. & Struhl, K. HBO1 histone acetylase activity is essential for DNA replication licensing and inhibited by Geminin. *Mol. Cell* **37**, 57–66 (2010).
28. Jacob, Y. *et al.* Regulation of heterochromatic DNA replication by histone H3 lysine 27 methyltransferases. *Nature* **466**, 987–991 (2010).
29. Brustel, J., Tardat, M., Kirsh, O., Grimaud, C. & Julien, E. Coupling mitosis to DNA replication: the emerging role of the histone H4-lysine 20 methyltransferase PR-Set7. *Trends Cell Biol.* **21**, 452–460 (2011).
30. Oda, H. *et al.* Monomethylation of histone H4-lysine 20 is involved in chromosome structure and stability and is essential for mouse development. *Mol. Cell Biol.* **29**, 2278–2295 (2009).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank R. Tennen for critical reading of the manuscript. This work was supported in part by grants to O.G. (R01 GM079641), D.J.P. (Abby Rockefeller Mauze, STARR and Maloris Foundations) and J.K.C. (DP1 OD003792), and a predoctoral fellowship to A.J.K. (Genentech Foundation). O.G. is a recipient of an Ellison Senior Scholar in Aging Award.

**Author Contributions** A.J.K. and P.C. performed the molecular biology, cellular and zebrafish studies; J.S. performed structural and binding affinity studies; S.Y. and J.K.C. advised on zebrafish experiments; S.I.-M. assisted in protein production and crystallization. A.J.K., P.C., J.S., D.J.P. and O.G. designed studies, analysed data and wrote the paper. All authors discussed and commented on the manuscript.

**Author Information** Atomic coordinates have been deposited with the Protein Data Bank under accession codes 4DOV (free structure) and 4DOW (H4K20me2-bound mORC1<sub>BAH</sub>) for the reported crystal structures. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to D.J.P. (pateld@mskcc.org) or O.G. (ogozani@stanford.edu).



## METHODS

**Materials and plasmids.** Biotinylated peptides were synthesized at Stanford or Yale Protein and Nucleic Acid facilities as previously described<sup>11</sup>. Antibodies used in this study: anti-histone H3 (Abcam), anti-Flag M5 (Sigma), anti-GST (Abcam), anti-tubulin (Upstate), anti-ORC2 (Upstate), anti-ORC3 (Abcam), anti-ORC5 (Sigma), anti-ORC6 (Abcam), anti-H4 (Abcam), anti-H4K20me1 (Abcam), anti-H4K20me2 (Abcam), anti-H4K20me3 (Abcam). ORC1 BAH domain was cloned into pGEX6P-1 for *in vitro* binding experiments, pBABE-puro-3×Flag for generating cell lines stably expressing ORC1 proteins, pCDNA-HA-EGFP and pCAG-Flag ORC1 for transfection in WI-38 cells, pCDNA for producing *in vitro*-transcribed mRNA for zebrafish microinjection. Site-directed mutagenesis was performed to introduce point mutations (Stratagene).

**Cell culture, cell synchronization and transfection.** 293T, HT1080 and U2OS cells were cultured in DMEM (Invitrogen) supplemented with 10% fetal bovine serum (ATCC), glutamine (Invitrogen) and penicillin/streptomycin (Invitrogen). U2OS cells were synchronized in 0.1 µg ml<sup>-1</sup> nocodazole (Sigma) for 24 h. Cells were collected for ChIP analysis and flow cytometry 6 h after releasing with DMEM. WI-38 human primary fibroblasts were maintained in DMEM/F12 (Invitrogen) with added supplements. Transient transfection was performed using TransIT-LT1 or TransIT-293 (Mirus) following the manufacturer's protocol.

**Flow cytometry.** Cells were collected and fixed in 70% ethanol at -20 °C for 15 min and rehydrated in PBS for 15 min. Cells were then stained with 3 µM propidium iodide and 10 µg ml<sup>-1</sup> RNase A (Sigma) in staining buffer (100 mM Tris, pH 7.4, 150 mM NaCl, 1 mM CaCl<sub>2</sub>, 0.5 mM MgCl<sub>2</sub>, 0.1% NP-40) for 30 min.

**Modified histone peptide microarrays.** Peptide microarray experiments were performed as described previously<sup>11</sup>. Briefly, biotinylated histone peptides diluted in phosphate-buffered saline (PBS) supplied with 5% glycerol and 0.05% Tween-20 were printed on streptavidin-coated glass slides (ArrayIt) using VersaArray Compact Microarrayer (BioRad). The arrays were probed with 25 µg of GST-ORC1<sub>BAH</sub> and protein-peptide interactions were detected using anti-GST antibodies (Abcam) followed by Alexa Fluor 647 chicken anti-rabbit IgG antibodies (Invitrogen). To validate the epitope-specificity of H4K20me2 antibodies, the arrays were probed with 1 µg of the indicated antibodies and the epitope-antibody bindings were detected by the same fluorophore-conjugated antibody as described above.

**Biotinylated peptide binding assay.** Biotinylated peptide pull-down assay was previously described<sup>31</sup>. Briefly, 1 µg of peptide was incubated with 1 µg of recombinant proteins in the binding buffer containing 50 mM Tris, pH 7.5, 150 mM NaCl and 0.05% NP-40. Peptides were pulled down using streptavidin sepharose beads (Amersham) and protein-peptide bindings were detected by western analyses.

**Full-length histone binding assay.** Full-length histone pull-down assays were performed as previously described<sup>32</sup>. Briefly, 5 µg of full-length histones purified from calf thymus (Wormington) was incubated with 25 µg of recombinant proteins in the binding solution containing 50 mM Tris, pH 7.5, 150 mM NaCl and 0.05% NP-40. Recombinant proteins were pulled down using glutathione sepharose 4B beads (Amersham) and bound histones were detected by western blot analysis.

**Protein expression and purification for structure analyses.** For structure determination, the gene encoding residues 9–170 of mouse ORC1 was PCR amplified and inserted into a modified pRSFDuet-1 vector (Novagen), in which mORC1<sub>BAH</sub> was separated from the preceding His<sub>6</sub>-SUMO tag by a ubiquitin-like protease (ULP1) cleavage site. The fusion protein was expressed in BL21 (DE3) RIL cell strain (Stratagene), in LB medium or minimum medium supplemented with 70 mg l<sup>-1</sup> seleno-methionine for selenium labelling. The cells were grown at 37 °C and induced by 0.4 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) when OD<sub>600 nm</sub> reached approximately 0.6. The temperature was then shifted to 20 °C for overnight culture growth. The fusion protein was purified through a Ni-NTA affinity column. The His<sub>6</sub>-SUMO tag was removed by ULP1 cleavage, followed by a second round of Ni-NTA column purification, and gel filtration on a 16/60 G200 Superdex column. The final sample for crystallization of the mORC1<sub>BAH</sub> domain contains about 10 mg ml<sup>-1</sup> mORC1<sub>BAH</sub> domain, 20 mM Tris-HCl, 50 mM NaCl, 5 mM DTT, pH 7.5.

For mutational analysis, mutants of the BAH domain human ORC1(1–185) were constructed by direct mutagenesis and purified as described above.

**Crystallization conditions.** For crystallization, the mORC1<sub>BAH</sub> domain was mixed with H4(14–25)K20me2 peptide in a molar ratio of 1:4. The crystallization condition (0.2 M sodium bromide, 25% PEG3350, 20 °C) was initially identified using sparse-matrix screens (Hampton Research). The crystals were subsequently reproduced and improved by the hanging-drop vapour-diffusion method, from drops mixed from 1 µl of H4K20me2-mORC1<sub>BAH</sub> domain solution and 1 µl of precipitant solution. This crystallization condition, when further supplemented with 0.2 M 3-(1-pyridino)-1-propane sulphonate (Hampton Research), led to production of the crystals of free mORC1<sub>BAH</sub>, even though the H4(14–25)K20me2 peptide was present in the crystallization solution. The crystals for both free and H4(14–25)K20me2-bound mORC1<sub>BAH</sub> domain were soaked in

cryoprotectant made of mother liquor supplemented with 25% glycerol, before flash freezing in liquid nitrogen.

**Structure determination.** X-ray diffraction data sets for both free and H4(14–25)K20me2-bound mORC1<sub>BAH</sub> domain were collected at selenium peak wavelength on the 24-ID-E-CAT beamline at the Advanced Photo Source, Argonne National Laboratory. The diffraction data were indexed, integrated and scaled using the HKL 2000 program. The structure of free mORC1<sub>BAH</sub> was solved by the single-wavelength dispersion method with selenium atoms using the AutoSol program embedded in PHENIX software<sup>33</sup>, which also gave an initial structural model. Further modelling of mORC1<sub>BAH</sub> domain was carried out using COOT<sup>34</sup>, and was then subject to refinement using PHENIX. The final model was refined to 1.70 Å resolution.

The structure of H4K20me2(14–25)-bound mORC1<sub>BAH</sub> domain was solved by the molecular replacement method in PHASER<sup>35</sup> using the free structure of the mORC1<sub>BAH</sub> domain as a search model. The H4(14–25)K20me2 peptide was then modelled in COOT and the structure of the H4(14–25)K20me2-mORC1<sub>BAH</sub> domain was refined using PHENIX. The final model of the complex was refined to 1.95 Å resolution. For both free and H4(14–25)K20me2-mORC1<sub>BAH</sub> domain structures, the *B* factors were refined with individual *B* values.

The statistics for data collection and structural refinement for both free and H4(14–25)K20me2-bound mORC1<sub>BAH</sub> domain are summarized in Supplementary Table 1.

**ITC measurements.** Protein and peptide samples used for ITC measurements were subject to overnight dialysis against buffer containing 20 mM Tris-HCl, 100 mM NaCl, 2 mM DTT, pH 7.5. Before the measurement, the protein and peptide concentrations were adjusted to about 0.1 mM and 1 mM, respectively. The ITC experiment was carried out using a MicroCal iTC200 instrument at 5 °C. The titration curves were analysed using software Origin7.0 (MicroCal, iTC200).

**Small-scale biochemical fractionation.** Small-scale biochemical fractionation was modified from a protocol described previously<sup>36</sup>. In short, 1 × 10<sup>7</sup> to 2 × 10<sup>7</sup> U2OS and HT1080 cells were collected, washed with PBS, and resuspended in buffer A (10 mM HEPES, pH 7.9, 10 mM KCl, 1.5 mM MgCl<sub>2</sub>, 0.34 M sucrose, 10% glycerol, 1 mM DTT, complete protease inhibitor tablet (Roche)). Triton X-100 was added to a final concentration of 0.1%. Cells were incubated for 8 min, and nuclei were collected by centrifugation (1,300g, 4 °C, 5 min). The supernatant (S1) was clarified by centrifugation at 20,000g, 4 °C for 5 min. The nuclei were washed once with buffer A and lysed in buffer B (3 mM EDTA, 0.2 mM EGTA, 1 mM DTT, complete protease inhibitor tablet) for 30 min. Chromatin was collected by centrifugation (1,700g, 4 °C, 5 min). The soluble fraction was combined with S1 and boiled at 90 °C with SDS sample buffer. Chromatin was washed with buffer B once and resuspended in SDS sample buffer, boiled at 90 °C for 10 min.

**Immunoprecipitation.** HT1080 and U2OS cells stably expressing Flag-ORC1 or Flag-ORC1 mutants were generated by retroviral transduction. Cells were lysed in cell lysis buffer containing 50 mM Tris-HCl, pH 7.4, 250 mM NaCl, 0.5% Triton X-100, 10% glycerol, 1 mM DTT, complete protease inhibitor tablet (Roche). Flag-ORC1 complexes were affinity purified by incubating anti-Flag m2 monoclonal antibody-conjugated agarose beads (Sigma) in lysates overnight at 4 °C. Flag M2 beads were then washed three times with cell lysis buffer and bound protein eluted in SDS buffer for western blot analysis.

**ChIP analysis.** ChIP was performed as previously described<sup>37</sup>. The primer sequences used in this study were: MCM4, -5 kbp forward, TTCACATCCA CCCAGCTTATC; MCM4, -5 kbp reverse, AGAGCATCTCTCCCTGATG; MCM4 origin, reverse, TTGGGTGGCTACTTGGTGT; MCM4 origin, reverse, TAGGCCCCCTCGCTTGT; MCM4, +5 kbp forward, TTTTGAATCTT GGTTTTGCTGA; MCM4, +5 kbp reverse, CAGCTTTTGGTTGGCTAAGG; β-globin, -40 kbp forward, AGGTCAGGCCCTCAAGAGTC; β-globin, -40 kbp reverse, CAGTCAGTTCTTTGGACAAGTCTTA; β-globin origin, forward, TGAGCCTCTGCTGATTCATTT; β-globin origin, reverse, TTCAAGGGAGAG ACCTCATGT; β-globin, +70 kbp forward, CTCAGAAGAACCCTTGAT CTCC; β-globin, +70 kbp reverse, TTCCGAGGACATTGGTTGA.

**Zebrafish aquaculture and microinjection.** Wild-type AB zebrafish (*D. rerio*) were maintained and raised using standard protocols. All experiments were conducted in accordance with AAALAC approved guidelines at Stanford University (protocol number 10511). Embryos obtained from natural matings were microinjected at the 1-to-2-cell stage with 1.5 nl of 0.18 mM sequence-specific *orc1* morpholino or 0.36 mM *suva420h1/h2* combined morpholinos at 1:1 ratio targeting each gene's translation start site (GeneTools). For epistasis analyses, 1.5 nl of morpholino solution containing 0.18 mM *orc1* morpholino and 0.36 mM *suva420h1/h2* combined morpholino (1:1 ratio) was used for microinjection. *orc1*, TCAGTCTTGATGTAGCGGCTCAT (described in ref. 4); *suva420h1*, ACCATGTCTTGGATTCTCCCATCT; *suva420h2*, CACTCATTCTATAAGT CCTCCAT. The morpholino targeting the *orc1* splice-site junction: ACAAC TCTATTATGCTACCTGTAC, as previously described<sup>4</sup>. For reconstitution



experiments, capped full-length ORC1 wild-type or H4K20me2-binding-pocket mutant mRNAs were transcribed *in vitro* following the manufacturer's manual (Ambion). Ten picograms of mRNA was co-microinjected with *orc1* morpholino at the 1-cell stage. Embryos were cultured in E3 medium at 28.5 °C for 24 h before scoring phenotypes. Live fish images were captured for individual fish using a Leica M205FA stereomicroscope and body length was calculated by Leica Application Suite program. Percentage of dwarf fish (average for control – 3× standard deviation of control) in each experiment group was scored and *P* values between groups were calculated by unpaired two-tailed Student's *t*-test.

31. Shi, X. *et al.* ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature* **442**, 96–99 (2006).
32. Matthews, A. G. *et al.* RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. *Nature* **450**, 1106–1110 (2007).
33. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
34. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
35. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
36. Mendez, J. & Stillman, B. Chromatin association of human origin recognition complex, Cdc6, and minichromosome maintenance proteins during the cell cycle: assembly of prereplication complexes in late mitosis. *Mol. Cell. Biol.* **20**, 8602–8612 (2000).
37. Kuo, A. J. *et al.* NSD2 links dimethylation of histone H3 at lysine 36 to oncogenic programming. *Mol. Cell* **44**, 609–620 (2011).

# MAP and kinesin-dependent nuclear positioning is required for skeletal muscle function

Thomas Metzger<sup>1,2\*</sup>, Vincent Gache<sup>3\*</sup>, Mu Xu<sup>1</sup>, Bruno Cadot<sup>3</sup>, Eric S. Folker<sup>1</sup>, Brian E. Richardson<sup>1</sup>, Edgar R. Gomes<sup>3,4\*</sup> & Mary K. Baylies<sup>1,2\*</sup>

The basic unit of skeletal muscle in all metazoans is the multi-nucleate myofibre, within which individual nuclei are regularly positioned<sup>1</sup>. The molecular machinery responsible for myonuclear positioning is not known. Improperly positioned nuclei are a hallmark of numerous diseases of muscle<sup>2</sup>, including centronuclear myopathies<sup>3</sup>, but it is unclear whether correct nuclear positioning is necessary for muscle function. Here we identify the microtubule-associated protein ensconsin (Ens)/microtubule-associated protein 7 (MAP7) and kinesin heavy chain (Khc)/Kif5b as essential, evolutionarily conserved regulators of myonuclear positioning in *Drosophila* and cultured mammalian myotubes. We find that these proteins interact physically and that expression of the Kif5b motor domain fused to the MAP7 microtubule-binding domain rescues nuclear positioning defects in MAP7-depleted cells. This suggests that MAP7 links Kif5b to the microtubule cytoskeleton to promote nuclear positioning. Finally, we show that myonuclear positioning is physiologically important. *Drosophila ens* mutant larvae have decreased locomotion and incorrect myonuclear positioning, and these phenotypes are rescued by muscle-specific expression of Ens. We conclude that improper nuclear positioning contributes to muscle dysfunction in a cell-autonomous fashion.

To identify the mechanisms governing nuclear positioning in muscle, we performed an F<sub>3</sub> recessive ethylmethane sulphonate mutagenesis screen using the *apterous*ME-NLS::dsRed transgenic line (*apRed*)<sup>4</sup>, which expresses dsRed in the nuclei of the four lateral transverse muscles of the *Drosophila* embryo. At the end of embryogenesis, the myonuclei are distributed throughout the muscles (Fig. 1a, b). One mutation resulted in clustered nuclei near the ventral end of each lateral transverse muscle. Within each hemisegment these clusters resembled the Nike trademark, so we named the mutant *swoosh* (*swo*) (Fig. 1a).

Further analysis of *swo* mutant embryos indicated that myoblast specification<sup>5</sup>, fusion<sup>5</sup>, muscle elongation<sup>6,7</sup> and attachment<sup>7</sup> occur normally. The only notable difference in *swo* mutant muscles is a ventral bulge in the cell body that correlates with the cluster of nuclei (Supplementary Figs 1 and 2). These data suggest that the clustering of nuclei in *swo* mutants is a genuine nuclear positioning defect.

To examine nuclear movement, we performed *in vivo* time-lapse imaging on *apRed* control and *swo, apRed* mutant embryos (Fig. 1b). In stage 14 *apRed* control embryos (10.5 h after egg laying (AEL)), nuclei in the four lateral transverse muscles group near the ventral end of each growing myotube in the hemisegment. During stage 15 (11.5 h AEL) after the completion of myoblast fusion and muscle attachment, the nuclei within each myotube begin to separate into distinct dorsal and ventral clusters. These clusters finish their migration during stage 16 (15 h AEL) and each cluster resides at an opposite end of the myotube. Finally, during stage 17 (18 h AEL), the six to eight nuclei in each lateral transverse muscle move from the clusters and distribute evenly throughout each lateral transverse muscle. The entire process of nuclear movement spans 7 h (Supplementary Movie 1). Time-lapse

imaging of *swo, apRed* mutant embryos demonstrated that the nuclei fail to undergo the initial separation into dorsal and ventral clusters during stage 15 and remain clustered at stage 17, indicating that *swo* does not simply cause a delay in myonuclear positioning (Supplementary Movie 2). In addition, the nuclei in every other somatic muscle examined in *swo* mutant embryos also failed to separate (Supplementary Fig. 1e).

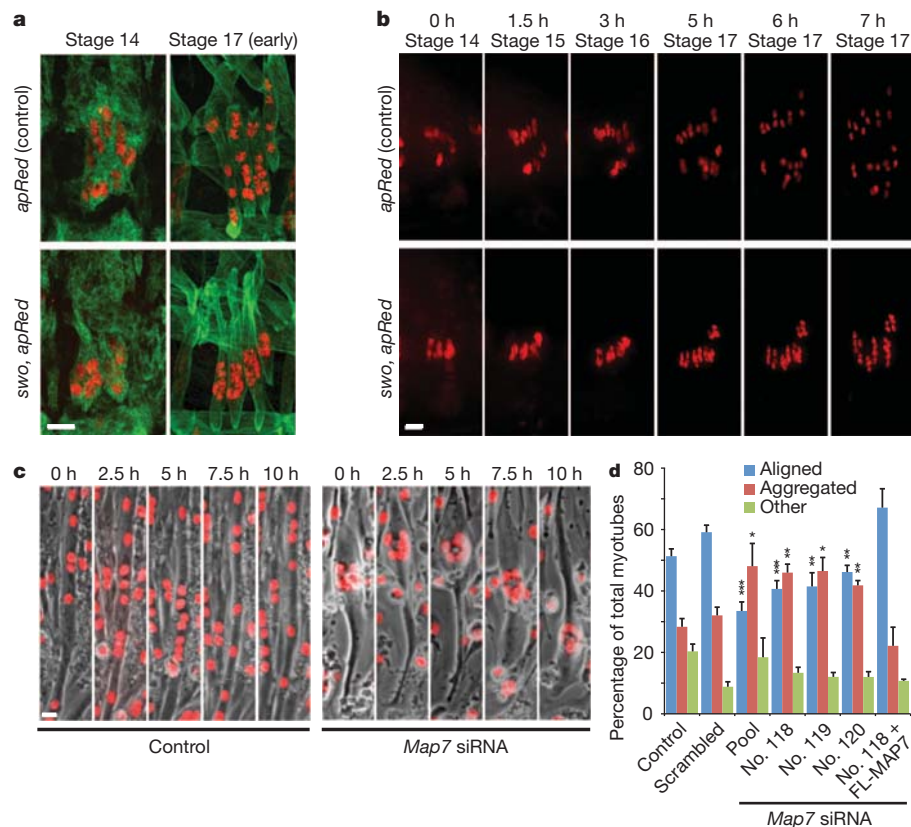
We determined that the *swo* flies carry a nonsense mutation in *ensconsin* (*ens*), a gene encoding a MAP. Expression of haemagglutinin-tagged Ens specifically in the developing mesoderm and muscle rescued the nuclear positioning phenotype in *swo* embryos, confirming that the mutation in *ens* is responsible for the nuclear positioning defect. Depletion of *ens* solely in the developing mesoderm and muscle by RNA-mediated interference recapitulated the phenotype we observed in the *swo* mutant (renamed *ens<sup>swo</sup>*), and known *ens* mutant alleles failed to complement *ens<sup>swo</sup>* (Supplementary Figs 1c and 2a, and data not shown). These results indicate that *ens* is required autonomously within muscle for proper myonuclear positioning. We also found that nuclear position was disrupted in *ens<sup>swo</sup>* oocytes<sup>8</sup> (Supplementary Fig. 3a, b), but not in *ens<sup>swo</sup>* photoreceptor cells (Supplementary Fig. 3c, and data not shown)<sup>9</sup>, indicating that Ens is necessary for some, but not all, nuclear positioning processes in *Drosophila*.

Four different genes in the mouse genome encode Ens orthologues: *Map7* (*E-MAP-115*, *Ensconsin*), *Map7D1*, *Map7D2* and *Map7D3* (Supplementary Fig. 2c). Using cultures of C2C12 myoblasts and primary mouse myoblasts, we examined the effect on nuclear positioning in myotubes depleted of each Ens orthologue by short interfering RNA (siRNA) (Fig. 1 and Supplementary Fig. 4). MAP7 depletion caused a significant increase in the aggregation of nuclei within a myotube but did not affect myoblast fusion or myotube differentiation (Fig. 1c, d, Supplementary Fig. 4d, e and Supplementary Movies 3–6). Depletion of MAP7D1, MAP7D2 or MAP7D3 did not affect nuclear positioning (Supplementary Fig. 4b). Expression of full-length MAP7 in MAP7-depleted myotubes restored nuclear alignment (Fig. 1d and Supplementary Fig. 6h). Thus, MAP7 is required for nuclear positioning in both *Drosophila* muscles and cultured mammalian myotubes.

No major defects on the microtubule network were observed in *ens<sup>swo</sup>* mutant embryos or MAP7-depleted myotubes (Supplementary Figs 2g–h and 6g). Therefore, to gain mechanistic insight, we conducted a yeast two-hybrid screen to find Ens-binding proteins and identified kinesin heavy chain (Khc) (Supplementary Fig. 5). Both a *kinesin*-null mutation (*khc<sup>8</sup>*) and a motor-dead mutation (*khc<sup>4</sup>*)<sup>10</sup> disrupted myonuclear positioning without affecting muscle elongation and attachment (Fig. 2a and Supplementary Fig. 5a). Likewise, siRNA-mediated depletion of the mammalian Khc orthologue Kif5b (ref. 11) in primary and C2C12 myotubes disrupted nuclear alignment similarly to MAP7 depletion without affecting myotube formation or differentiation (Fig. 2b, c, Supplementary Figs 4d–e

<sup>1</sup>Program in Developmental Biology, Sloan-Kettering Institute, New York, New York 10065, USA. <sup>2</sup>Weill Graduate School of Medical Sciences of Cornell University, New York, New York 10065, USA. <sup>3</sup>UMR S 787 INSERM, Université Pierre et Marie Curie Paris 6, 75634 Paris, France. <sup>4</sup>Groupe Hospitalier Pitié-Salpêtrière, Institut de Myologie, 75013 Paris, France.

\*These authors contributed equally to this work.

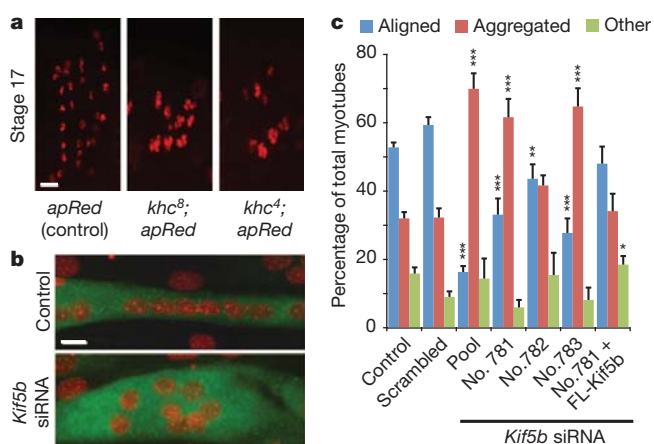


**Figure 1 | Myonuclear positioning requires ensconsin/MAP7.**

**a**, Hemisegments from control and *ens<sup>sws</sup>* embryos at the beginning (stage 14) and end (stage 17) of nuclear migration. Muscles are immunostained for tropomyosin (green); nuclei are immunostained for dsRed (red). Scale bar, 10  $\mu$ m. **b**, Time-lapse images of nuclear migration in hemisegments of control and *ens<sup>sws</sup>* embryos. Elapsed imaging times are shown along the top. Scale bar, 10  $\mu$ m. **c**, Time-lapse images of nuclear migration in control and *Map7*-

depleted C2C12-H1B-GFP myotubes. Elapsed imaging times are shown along the top. Nuclei are stained red. Scale bar, 15  $\mu$ m. **d**, Nuclear distribution in C2C12 myotubes that were untreated (control) or treated with the indicated siRNA and *Map7*-depleted cells expressing full-length MAP7 (no. 118 + FL-MAP7). Error bars indicate s.e.m. Asterisk,  $P < 0.05$ ; two asterisks,  $P < 0.01$ ; three asterisks,  $P < 0.001$  (scrambled versus experimental condition).

depleted C2C12-H1B-GFP myotubes. Elapsed imaging times are shown along the top. Nuclei are stained red. Scale bar, 15  $\mu$ m. **d**, Nuclear distribution in C2C12 myotubes that were untreated (control) or treated with the indicated siRNA and *Map7*-depleted cells expressing full-length MAP7 (no. 118 + FL-MAP7). Error bars indicate s.e.m. Asterisk,  $P < 0.05$ ; two asterisks,  $P < 0.01$ ; three asterisks,  $P < 0.001$  (scrambled versus experimental condition).



**Figure 2 | Kinesin is required for myonuclear positioning.** **a**, Single hemisegments from stage 17 (22 h AEL) embryos of indicated genotypes. Scale bar, 10  $\mu$ m. **b**, Representative immunofluorescence images of control and *Kif5b*-depleted C2C12 myotubes differentiated for 4 days and immunostained for myosin heavy chain (green) and 4',6-diamidino-2-phenylindole (red). Scale bar, 15  $\mu$ m. **c**, Histogram of nuclear distribution in C2C12 myotubes that were untreated (control) or treated with the indicated siRNA or *Kif5b*-depleted cells expressing full-length Kif5b (no. 781 + FL-Kif5b). Error bars indicate s.e.m. Asterisk,  $P < 0.05$ ; two asterisks,  $P < 0.01$ ; three asterisks,  $P < 0.001$  (scrambled versus experimental condition).

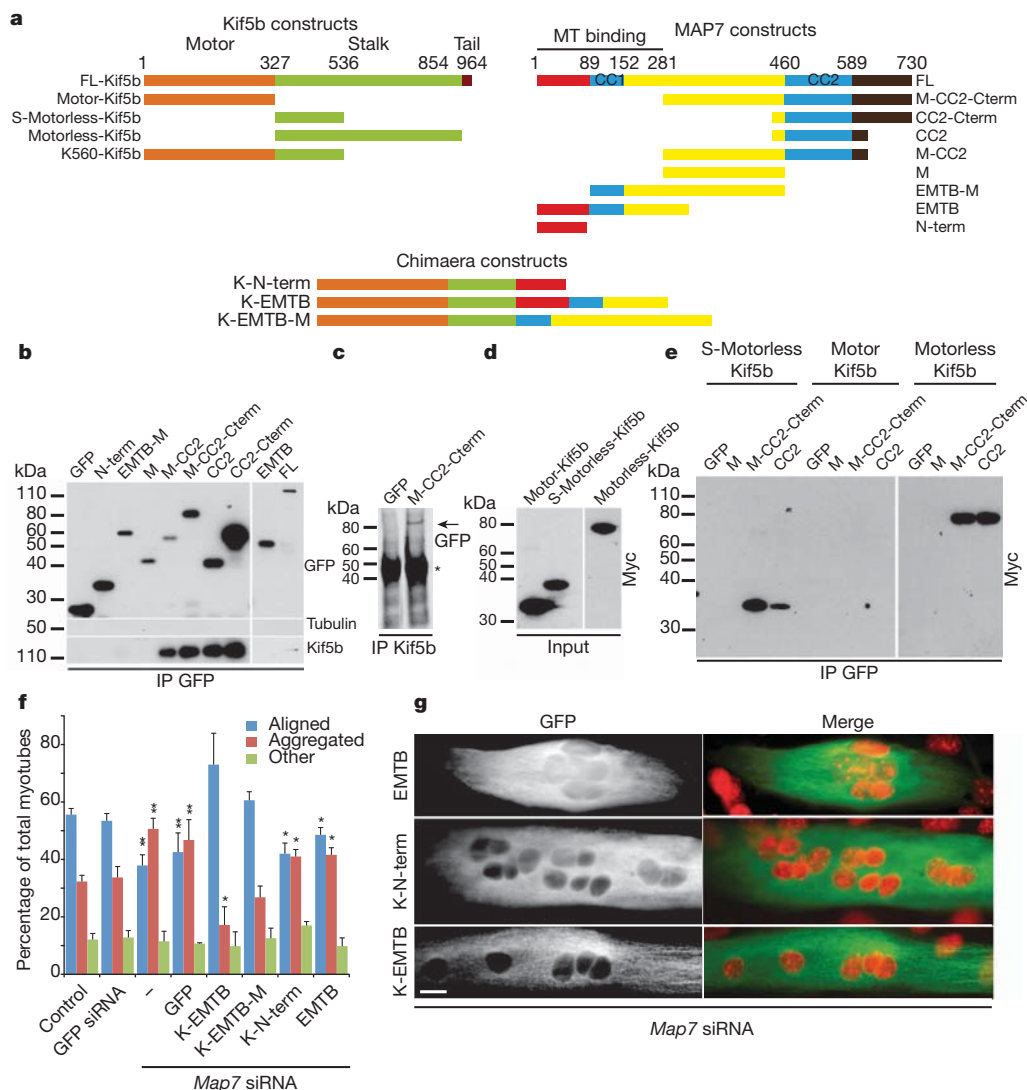
functional motor domain, is required for proper nuclear positioning in both *Drosophila* muscles and cultured mammalian myotubes.

We next examined whether Ens and Khc interact functionally to position myonuclei. Embryos heterozygous for *ens* or *khc* position their myonuclei normally. However, embryos that were doubly heterozygous for *ens<sup>sws</sup>* and *khc* showed defective nuclear positioning similar to that in the homozygous *ens* embryos (Supplementary Fig. 5). This demonstrates that *ens* and *khc* interact functionally and that the dosage of these genes is important for proper nuclear positioning.

We next tested whether Ens/MAP7 and Khc/Kif5b interact physically. Endogenous Kif5b co-immunoprecipitated with expressed full-length green fluorescent protein (GFP)-tagged MAP7, independently of microtubules (Fig. 3a, b). Using fragments of MAP7, we mapped the Kif5b interaction to the carboxy-terminal coiled-coil domain of MAP7 (CC2), which was confirmed by reciprocal co-immunoprecipitation (Fig. 3a–c). Similarly, fragments of Kif5b were used to identify the C-terminal region (Kif5b-motorless) as the MAP7-interacting domain (Fig. 3a, d, e). The C-terminal coiled-coil domain of MAP7 therefore interacts with the C-terminal region of Kif5b (Fig. 3).

To test whether the specific interaction identified above between Ens/MAP7 and Khc/Kif5b is required for nuclear positioning, we expressed chimaeras consisting of the Kif5b motor domain<sup>12</sup> fused to fragments of MAP7 that contained (K-EMTB and K-EMTB-M) or lacked (K-N-term) the microtubule-binding domain (EMTB)<sup>13</sup> in *Map7* siRNA-depleted myotubes (Fig. 3a, f, g). Expression of chimaeras that contained the MAP7 EMTB rescued the alignment of nuclei in MAP7-depleted myotubes, whereas expression of the EMTB domain alone, or a chimaera that lacked the MAP7 EMTB (K-N-term) did not





**Figure 3 | Kinesin and ensconsin/MAP7 interact to regulate nuclear position.** **a**, Kif5b, MAP7 and chimaera constructs amino-terminally tagged with GFP (MAP7 and chimaeras) or c-Myc (Kif5b). Numbers indicate amino acids. **b**, Western blot with indicated antibodies (right) of GFP-MAP7 (top) immunoprecipitations (IP). **c**, Anti-GFP Western blot of Kif5b immunoprecipitations from C2C12 cells expressing indicated constructs (top). Asterisk indicates anti-Kif5b IgG. **d**, Anti-c-Myc Western blot from C2C12 cells

expressing Kif5b constructs. **e**, Anti-c-Myc Western blot from GFP immunoprecipitations with C2C12 myotubes expressing the indicated constructs. **f**, Nuclear distribution in *Map7*-depleted C2C12 myotubes expressing the indicated chimaeras. Error bars indicate s.e.m. Asterisk,  $P < 0.05$ ; two asterisks,  $P < 0.01$ . **g**, *Map7*-depleted C2C12s expressing indicated chimaeras. Anti-GFP is shown in green; nuclei are shown in red (stained with 4',6-diamidino-2-phenylindole). Scale bar, 15 μm.

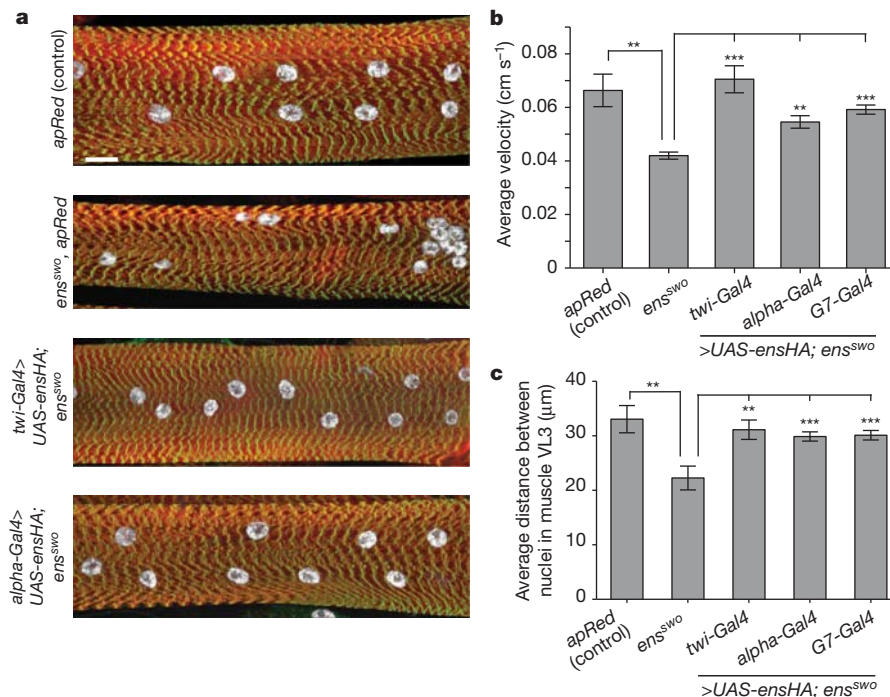
(Fig. 3f, g). These results indicate that the function of MAP7 in nuclear positioning is to link Kif5b to microtubules.

To investigate the physiological impact of mispositioned nuclei on muscle function *in vivo*, we examined larval motility<sup>14</sup>. *ens* mutant larvae move significantly more slowly than controls (Fig. 4b and Supplementary Fig. 7d). Moreover, analysis of the muscle structure in tracked larvae revealed that the myonuclei in *ens*<sup>sw</sup> mutants were 33% closer together than in controls (Fig. 4a, c and Supplementary Fig. 7a, e). Sarcomere structure<sup>15</sup>, t-tubules<sup>16</sup>, neuromuscular junctions<sup>17</sup>, nuclei number (Supplementary Fig. 7c, e, f, h, i) and mitochondrial localization (Supplementary Fig. 8) were unchanged in *ens* mutants compared with controls. Expression of *Ens* in the mesoderm and muscle during embryonic development or in muscle starting at the L2 larval stage in *ens* mutants rescued the nuclear positioning and motility defects in L3 larvae (Fig. 4a–c and Supplementary Fig. 7a, d, g), demonstrating a strong correlation between aberrant myonuclear position and decreased muscle function.

These experiments define a subset of the nuclear behaviours that occur during muscle differentiation, and demonstrate conservation

across species at both the cellular and molecular levels. We find that *Ens*/MAP7 and *Khc*/Kif5b are critical for myonuclear positioning and are not required for fusion and myofibre formation. Moreover, these proteins interact genetically and physically, and the physical interaction is necessary for proper nuclear positioning. Furthermore, our results argue that the correct spacing of myonuclei is required for proper muscle function.

Within a myotube, distinct microtubule networks emanate from each myonucleus, producing regions of overlapping antiparallel microtubules<sup>18–20</sup>. Furthermore, *Ens*/MAP7 functions to load *Khc*/Kif5b onto microtubules<sup>8</sup>. The results presented here therefore suggest a molecular mechanism for nuclear positioning. Adjacent nuclei could be positioned relative to each other by *Ens*/MAP7 and *Khc*/Kif5b interacting with and sliding antiparallel microtubules that have their minus ends anchored to the nuclear envelope, in a similar manner to the mechanism by which the Eg5 kinesin facilitates spindle elongation<sup>21</sup> (Supplementary Fig. 10). This mechanism is distinct from the kinesin-1-dependent nuclear position in the *Caenorhabditis elegans* hypodermis, in which *Khc* is anchored to the nucleus by kinesin light



**Figure 4 | Ensconsin/Map7 is required for intracellular muscle organization and efficient larval locomotion.** **a**, Maximum-intensity xy projections of muscle VL3 from segment A3 of L3 larvae from the indicated genotypes stained for actin (red), nuclei (white) and Z-bands (green). Scale bar, 20 μm. **b**, Average velocity of migration for L3 larvae of the indicated genotypes.

chain (Klc) and the KASH-domain protein Unc-83, which drive the nucleus unidirectionally on microtubules<sup>22</sup>. We find that *klc* mutants in *Drosophila* do not show defects in the positioning of nuclei that we describe here. Additionally, overexpression of dominant-negative (dn)-KASH-domain constructs in C2C12 myotubes that remove KASH proteins from the nuclear envelope and disrupt the function of the LINC (linker of nucleoskeleton and cytoskeleton) complex and nuclear anchoring<sup>23,24</sup> does not affect nuclear positioning (Supplementary Fig. 9). Taken together, these results suggest that kinesin is not interacting directly with the nucleus; they therefore provide evidence for the sliding mechanism.

Patients suffering from various muscle diseases, including centronuclear myopathies, show muscle weakness with mislocalized myonuclei<sup>3</sup>. However, it remains unclear whether these mispositioned nuclei contribute to muscle weakness or are simply a result of impaired myofibre function<sup>3</sup>. Our results show that the ability to position myonuclei correctly correlates with better muscle function. Because nuclear misposition is the first observable defect in zygotic *ens* mutant embryos, we speculate that this is a cause of the observed muscle weakness. Mispositioned nuclei might lead to the observed muscle weakness as a result of a disruption in the size and spacing of myonuclear domains throughout the muscle<sup>25</sup> or in the distribution of subcellular structures whose positioning depends on correct nuclear spacing. We rescue nuclear position, muscle function, and viability not only by supplying *Ens* during embryogenesis but also by expressing *Ens* solely in mature muscles later in larval development. We therefore propose that correcting nuclear positioning defects in patients with muscle diseases might benefit muscle strength and improve muscle function. The model systems that we present here provide a platform with which to identify additional nuclear positioning genes and assess their function, and to screen for potential drugs to alleviate muscle weakness due to disease.

## METHODS SUMMARY

*Drosophila* stocks used included *apME-NLS::dsRed* (ref. 4), *Df(3L)GN34* (ref. 26), *ens*<sup>HP36480</sup> (Bloomington), *ens*<sup>107121</sup> (Harvard), *Df-ens*<sup>A3277</sup>, *ens*<sup>AN</sup>, *ens*<sup>AC</sup> (ref. 8),

Each of the indicated Gal4 drivers is expressing *UAS-ensHA* in a homozygous *ens*<sup>swo</sup> mutant background. Error bars, s.e.m. Two asterisks,  $P < 0.01$ ; three asterisks,  $P < 0.001$ . **c**, Nearest-neighbour analysis of nuclei within muscle VL3 from segment A3 from L3 larvae of the indicated genotypes. Error bars indicate s.e.m. Asterisk,  $P < 0.05$ ; three asterisks,  $P < 0.001$ .

*khc*<sup>8</sup>, *khc*<sup>4</sup>, *khc*<sup>23</sup> (ref. 10), *klc*<sup>8ex94</sup> (ref. 27), *twi-Gal4* (ref. 4), *alpha-Gal4*, *G7-Gal4* (ref. 28) (from K. Broadie), and *UAS-ens-IR* lines 106207 and 18491 (VDR). *UAS-ensHA* transgenic flies were generated by BestGene Inc. Mouse cells were transfected with siRNA using Lipofectamine RNAiMAX (Invitrogen) or with DNA using Lipofectamine 2000 (Invitrogen). Primary antibodies used were rabbit anti-dsRed (Clontech), rat anti-tropomyosin (Abcam), mouse anti-myosin heavy chain (from S. Abmayr), rabbit anti-Zasp (from F. Schöck), chicken anti-β-gal (Abcam), mouse anti-α-tubulin (Sigma), rat anti-ensconsin (from P. Rørth), guinea pig anti-Krüppel (from J. Reinitz), rabbit anti-Eve (from M. Frasch), mouse anti-β-PS integrin (DSHB), rabbit anti-vestigial (from S. Carroll), fluorescein isothiocyanate-conjugated anti-horseradish peroxidase (Jackson ImmunoResearch), mouse anti-discs large (DSHB), rabbit anti-ATP synthase<sup>29</sup> (from H. Duan), rat anti-DE-cadherin (DSHB), mouse anti-chaoptin 24B10 (DSHB), rat anti-elav (DSHB), Alexa Fluor 488-conjugated wheatgerm agglutinin (Invitrogen), MF20 (DSHB), rabbit anti-KHC (Santa Cruz), mouse anti-c-Myc (Roche) and rabbit anti-GFP (Invitrogen). Secondary antibodies were either biotinylated (Vector Laboratories and Jackson ImmunoResearch) or conjugated to Alexa Fluor 488, 555 or 647. The fusion index<sup>30</sup>, sarcomere length<sup>15</sup>, bouton number<sup>17</sup> and larval velocity<sup>14</sup> were quantified as described, with minor modifications (see Methods). The yeast two-hybrid screen was performed with full-length *Ens* by Hybrigenics SA Services using a 0–24-h *Drosophila* complementary DNA library. Standard protocols were used for immunoprecipitation, western blotting and quantitative polymerase chain reaction experiments and are described in Methods.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 30 March 2011; accepted 31 January 2012.

Published online 18 March 2012.

1. Bruusgaard, J. C., Liestøl, K., Ekmark, M., Kollstad, K. & Gundersen, K. Number and spatial distribution of nuclei in the muscle fibres of normal mice studied *in vivo*. *J. Physiol. (Lond.)* **551**, 467–478 (2003).
2. Cohn, R. D. & Campbell, K. P. Molecular basis of muscular dystrophies. *Muscle Nerve* **23**, 1456–1471 (2000).
3. Jungbluth, H., Wallgren-Pettersson, C. & Laporte, J. Centronuclear (myotubular) myopathy. *Orphanet J. Rare Dis.* **3**, 26 (2008).
4. Richardson, B. E., Beckett, K., Nowak, S. J. & Baylies, M. K. SCAR/WAVE and Arp2/3 are crucial for cytoskeletal remodeling at the site of myoblast fusion. *Development* **134**, 4357–4367 (2007).

5. Beckett, K. & Baylies, M. K. The development of the *Drosophila* larval body wall muscles. *Int. Rev. Neurobiol.* **75**, 55–70 (2006).
6. Prokop, A., Martín-Bermudo, M. D., Bate, M. & Brown, N. H. Absence of PS integrins or laminin A affects extracellular adhesion, but not intracellular assembly, of hemiadherens and neuromuscular junctions in *Drosophila* embryos. *Dev. Biol.* **196**, 58–76 (1998).
7. Volk, T. Singling out *Drosophila* tendon cells: a dialogue between two distinct cell types. *Trends Genet.* **15**, 448–453 (1999).
8. Sung, H.-H. *et al.* *Drosophila* ensconsin promotes productive recruitment of Kinesin-1 to microtubules. *Dev. Cell* **15**, 866–876 (2008).
9. Fischer-Vize, J. A. & Mosley, K. L. Marbles mutants: uncoupling cell determination and nuclear migration in the developing *Drosophila* eye. *Development* **120**, 2609–2618 (1994).
10. Brenda, K. M., Rose, D. J., Gilbert, S. P. & Saxton, W. M. Lethal kinesin mutations reveal amino acids important for ATPase activation and structural coupling. *J. Biol. Chem.* **274**, 31506–31514 (1999).
11. Vale, R. D., Reese, T. S. & Sheetz, M. P. Identification of a novel force-generating protein, kinesin, involved in microtubule-based motility. *Cell* **42**, 39–50 (1985).
12. Vale, R. D. *et al.* Direct observation of single kinesin molecules moving along microtubules. *Nature* **380**, 451–453 (1996).
13. Masson, D. & Kreis, T. E. Identification and molecular characterization of E-MAP-115, a novel microtubule-associated protein predominantly expressed in epithelial cells. *J. Cell Biol.* **123**, 357–371 (1993).
14. Louis, M., Huber, T., Benton, R., Sakmar, T. P. & Vosshall, L. B. Bilateral olfactory sensory input enhances chemotaxis behavior. *Nature Neurosci.* **11**, 187–199 (2008).
15. Bai, J., Hartwig, J. H. & Perrimon, N. SALS, a WH2-domain-containing protein, promotes sarcomeric actin filament elongation from pointed ends during *Drosophila* muscle growth. *Dev. Cell* **13**, 828–842 (2007).
16. Razaq, A. *et al.* Amphiphysin is necessary for organization of the excitation-contraction coupling machinery of muscles, but not for synaptic vesicle endocytosis in *Drosophila*. *Genes Dev.* **15**, 2967–2979 (2001).
17. McCabe, B. D. *et al.* The BMP homolog Gbb provides a retrograde signal that regulates synaptic growth at the *Drosophila* neuromuscular junction. *Neuron* **39**, 241–254 (2003).
18. Tassin, A. M., Maro, B. & Bornens, M. Fate of microtubule-organizing centers during myogenesis *in vitro*. *J. Cell Biol.* **100**, 35–46 (1985).
19. Guerin, C. M. & Kramer, S. G. RacGAP50C directs perinuclear  $\gamma$ -tubulin localization to organize the uniform microtubule array required for *Drosophila* myotube extension. *Development* **136**, 1411–1421 (2009).
20. Bugnard, E., Zaal, K. J. M. & Ralston, E. Reorganization of microtubule nucleation during muscle differentiation. *Cell Motil. Cytoskeleton* **60**, 1–13 (2005).
21. Glotzer, M. The 3Ms of central spindle assembly: microtubules, motors and MAPs. *Nature Rev. Mol. Cell Biol.* **10**, 9–20 (2009).
22. Meyerzon, M., Fridolfsson, H. N., Ly, N., McNally, F. J. & Starr, D. A. UNC-83 is a nuclear-specific cargo adaptor for kinesin-1-mediated nuclear migration. *Development* **136**, 2725–2733 (2009).
23. Grady, R. M., Starr, D. A., Ackerman, G. L., Sanes, J. R. & Han, M. Syne proteins anchor muscle nuclei at the neuromuscular junction. *Proc. Natl Acad. Sci. USA* **102**, 4359–4364 (2005).
24. Lei, K. *et al.* SUN1 and SUN2 play critical but partially redundant roles in anchoring nuclei in skeletal muscle cells in mice. *Proc. Natl Acad. Sci. USA* **106**, 10207–10212 (2009).
25. Pavlath, G. K., Rich, K., Webster, S. G. & Blau, H. M. Localization of muscle gene products in nuclear domains. *Nature* **337**, 570–573 (1989).
26. Feng, G., Deák, P., Kasbekar, D. P., Gil, D. W. & Hall, L. M. Cytogenetic and molecular localization of tipE: a gene affecting sodium channels in *Drosophila melanogaster*. *Genetics* **139**, 1679–1688 (1995).
27. Gindhart, J. G., Desai, C. J., Beushausen, S., Zinn, K. & Goldstein, L. S. Kinesin light chains are essential for axonal transport in *Drosophila*. *J. Cell Biol.* **141**, 443–454 (1998).
28. Zhang, Y. Q. *et al.* *Drosophila* fragile X-related gene regulates the MAP1B homolog Futsch to control synaptic structure and function. *Cell* **107**, 591–603 (2001).
29. Peña, P. & Garesse, R. The  $\beta$  subunit of the *Drosophila melanogaster* ATP synthase: cDNA cloning, amino acid analysis and identification of the protein in adult flies. *Biochem. Biophys. Res. Commun.* **195**, 785–791 (1993).
30. Mitchell, K. J. *et al.* Identification and characterization of a non-satellite cell muscle resident progenitor during postnatal development. *Nature Cell Biol.* **12**, 257–266 (2010).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank K. Anderson, K. Hadjantonakis, A. Hall and D. Sassoon for comments on the manuscript. We thank the Baylies and Gomes Laboratories for discussions, and R. Fernandez-Gonzalez for his assistance in computational analysis. The initial screens in *Drosophila* were supported by National Institutes of Health (NIH) grants GM056989 and GM0781318 to M.B.; the *Drosophila* nuclear positioning analysis was supported by a Muscular Dystrophy Association (MDA) grant to M.B. T.M. was supported initially by NIH Training Grant T32 BM008539. B.C. was supported initially by a Fondation pour la Recherche Médicale (FRM) fellowship. V.G. was supported initially by a Région Île-de-France fellowship. Mammalian work was supported by Muscular Dystrophy Association (MDA), INSERM Avenir programme and Agence Nationale de la Recherche (ANR) grants to E.R.G.

**Author Contributions** T.M. and V.G. are joint first authors. E.G. and M.B. are joint senior authors. M.B., M.X., T.M., E.G., V.G. and B.C. conceived, designed and analysed the experiments. T.M., M.X., E.S. and B.R. conducted the *Drosophila* experimental work. V.G. and B.C. conducted the mouse primary cultures and C2C12 experimental work. The manuscript was written by T.M., E.S., V.G., E.G. and M.B. with assistance from other authors.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence should be addressed to E.G. ([edgar.gomes@upmc.fr](mailto:edgar.gomes@upmc.fr)) or M.B. ([mb-baylies@ski.mskcc.org](mailto:mb-baylies@ski.mskcc.org)).



## METHODS

**Flies and husbandry.** Stocks were grown and maintained under standard conditions on standard cornmeal medium. Crosses were performed at 25 °C. *Drosophila* stocks: *apME-NLS::dsRed* (ref. 4), *Df(3L)GN34* (ref. 26), *ens*<sup>HP36480</sup> (Bloomington), *ens*<sup>RO7121</sup> (Harvard), *ens*<sup>AN</sup>, *ens*<sup>AC</sup>, *Df-ens*<sup>A3277</sup> (ref. 8), *UAS-ensHA* (this study), *khc*<sup>8</sup>, *khc*<sup>8</sup>, *khc*<sup>23</sup> (ref. 10), *klc*<sup>8ex94</sup> (ref. 27), and *UAS-ens-IR* lines 106207 and 18491 (VDR). Mutant alleles were balanced and identified using *CyO P[w+en11lacZ]* or *CTG [CyO, twi-GAL4, UAS-2xeGFP]* (ref. 26), *TM3 Sb1Dfd-lacZ* or *TTG [TM3, twi-GAL4, UAS-2xeGFP]* (ref. 31). Gal4 lines used in this work were *twi-Gal4* (ref. 4), which is specifically expressed in the mesoderm and developing muscle (2.5–13 h AEL), and *alpha-Gal4* and *G7-Gal4* (ref. 28) (gift from K. Broadie), which are specifically expressed in the larval muscles beginning at the late L1 and early L2 larval stages and perdures until pupation.

**Ethylmethane sulphonate mutagenesis and screening.** A detailed description of the screen and its findings are available from the authors. In brief, males from an isogenized stock carrying the *apME-NLS::dsRed* transgene on the third chromosome were mutagenized with 35 mM ethylmethane sulphonate dissolved in 1% sucrose in accordance with a standard protocol<sup>32</sup>. Individual mutant lines were established over the TTG balancer. Live F<sub>3</sub> progeny were screened under fluorescence, and embryos homozygous for the mutant chromosome were identified by an absence of the balancer GFP. Deficiency mapping, complementation and sequencing were used to identify the genetic lesion.

**Phylogenetic analysis.** *Drosophila*, human and mouse MAP7 orthologues were aligned using LALIGN ([http://www.ch.embnet.org/software/LALIGN\\_form.html](http://www.ch.embnet.org/software/LALIGN_form.html)) and the phylogenetic tree was created using Vector NTI (Invitrogen).

**Cell culture.** C2C12 myoblasts were grown and differentiated for 4 days as described<sup>33</sup>. Primary myoblasts were isolated and differentiated for 3 days as described<sup>30,34</sup>.

**Constructs.** *Map7* fragments were cloned from a C2C12 cDNA library into pEGFP-C1 vector (Invitrogen), using the following primers: *N-term* (1–244), 5'-ATGGCGGAGCAGGAGC-3' (forward) and 5'-CTAGCTGTTTCTCTGTTCTC-3' (reverse); *EMTB-M* (261–1351), 5'-ATGTGGCTAGAGAGAGAA GAACGAG-3' (forward) and 5'-TAGGAACACACCGACAGTCACAG-3' (reverse); *M-CC2-Cterm* (859–2187), 5'-ATGACCATTCATGGACTAGCG AG-3' (forward) and 5'-AACTTCTGCGGTCTGTGT-3' (reverse); *M-CC2* (859–1806), 5'-ATGACCATTCATGGACTAGCGAG-3' (forward) and 5'-TT TCTTATCAGCGGTCTCTGTCTC-3' (reverse); *CC2-Cterm* (1337–2187), 5'-ATGGGCGGTGTAGTTCCTAAGACTTCTG-3' (forward) and 5'-AACTT CTGCGGTCTGTGTGTG-3' (reverse); *CC2* (1337–1806), 5'-ATGGGCGGTGT AGTTCCTAAGACTTCTG-3' (forward) and 5'-TTTCTTATCAGCGGTCTCT GTCTC-3' (reverse); *M* (859–1351), 5'-ATGACCATTCATGGACTAGC GAG-3' (forward) and 5'-TAGGAACACACCGACAGTCACAG-3' (reverse).

Full-length *Map7* (NM\_008635) was made from the *Map7* fragments and was cloned into pEGFP-C1. *EMTB-Map7* was made by cloning the fragment consisting of base pairs 1–947 from full-length *Map7* into the pEGFP-C1 vector. *K-N-term*, *K-EMTB* and *K-EMTB-M* chimera constructs were made by cloning the K560 fragment from pET17b-K560-GFP (gift from R. Vale) into *N-Term*, *EMTB* and *EMTB-M* *Map7* pEGFP constructs, respectively. *Motor-Kif5b*, *S-Motorless-Kif5b* and *Motorless-Kif5b* c-Myc were a gift from G. Kreitzer. FL-Kif5b was made by cloning the K560 fragment with a C-terminal region of full-length Kif5b (gift from C. C. Hoogenraad) into pEGFP C1 construct. EGFP-KASH2 and EGFP-KASH2ext were a gift from D. Hodzic<sup>35</sup>.

siRNA sequences (Ambion): *Map7*, 5'-CAGAUUAGAUGUCACCAAUTT-3' (no. 118), 5'-CCAUGAAUCUUUCGAAACATT-3' (no. 119) and 5'-ACUUAUCUGUUGGAUCAAATT-3' (no. 120); *Map7D1*, 5'-GGAACAGAGGGAACG CGAATT-3' (no. 710), 5'-ACGUGGACUCUAUUAUCAAATT-3' (no. 711) and 5'-GCAAUCCAGCGGUCAGUGATT-3' (no. 712); *Map7D2*, 5'-CUUUAACAUCAACCAUGATT-3' (no. 436), 5'-CUUGAUGACUGUAACAAATT-3' (no. 437) and 5'-GGUCUCCUGUAAGUAUUAATT-3' (no. 438); *Map7D3*, 5'-GGACGACAUCUCUAAGUUTT-3' (no. 000), 5'-GAACUAUCCUCCAU AGUGATT-3' (no. 001) and 5'-CUGAUGAGGUUAUACCAATT-3' (no. 002); *Kif5b*, 5'-GCAAGAAGUAGACCGGAUATT-3' (no. 781), 5'-GCUUUAUUAUGAUCAGATT-3' (no. 782) and 5'-GACAUGUCGAGUUAACAAATT-3' (no. 783).

**Transfections.** Myoblasts were transfected with siRNA using Lipofectamine RNAiMAX (Invitrogen) or DNA using Lipofectamine 2000 (Invitrogen). H1B-GFP stable cell line was made by transfecting cells with H1B-GFP (gift from R. Vallee).

**qPCR.** Messenger RNA from C2C12 cells was isolated with the RNeasy Micro Kit (Qiagen), and cDNA was prepared with the Transcriptor first-strand cDNA synthesis kit (Roche). Quantitative analyses (qPCR) were performed using CyberGreen kit in a LightCycler 480 II system (Roche), using the following primers: *Kif5b-F*, 5'-GGAGGCAAGCAGTCGTAAAC-3'; *Kif5b-R*, 5'-TCTAG

TGTTGGGAAGCAGCA-3'; *Map7-F*, 5'-TGAAACCAATTTTGGCCTTT-3'; *Map7-R*, 5'-AAAACACAGAGGAGGAGCTG-3'; *Hprt1-F*, 5'-GTTAAGCAGTACAGCCCCAAA-3'; *Hprt1-R*, 5'-AGGGCATATCCAACAACAACTT-3'.

**Immunoprecipitations.** C2C12 cells were transfected with GFP-tagged or c-Myc-tagged constructs. After 24 h, cell lysates were prepared using lysis buffer (10 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.5 mM EDTA, 0.5% Nonidet P40, 1 mM phenylmethylsulphonyl fluoride) containing protease inhibitor cocktail (Roche) and 0.6 μM nocodazole (Sigma). Immunoprecipitation of GFP-fusion proteins or endogenous Kif5b was performed at 4 °C using the magnetic GFP-Trap kit (Chromotek) or Protein A magnetic beads (Millipore) coated with Khc antibody, respectively.

**Western blot.** Extracts were loaded on a 4–12% gradient gel (Invitrogen) and transferred using iBlot (Invitrogen).

**Immunohistochemistry, immunofluorescence, microscopy, and image analysis.** Embryos were collected at 25 °C on apple-juice agar plates and were fixed as described<sup>4</sup> except for tubulin staining, for which the embryos were fixed in equal volumes of 10% formalin (Sigma) and heptane (Fisher Scientific). Larvae were dissected and fixed in 4% paraformaldehyde (Electron Microscopy Sciences) in relaxing buffer as described previously<sup>15,36,37</sup>. Eye discs were dissected from wandering third-instar larvae and fixed in 4% paraformaldehyde. Ovaries from females that were fed on fresh yeast for 2 days were dissected in PBS and fixed in 4% paraformaldehyde. All fluorescent stainings were mounted in Prolong Gold (Molecular Probes); otherwise, embryos were mounted in Araldite.

Muscle cells were fixed using 3.7% paraformaldehyde (Sigma) and permeabilized with 0.5% Triton. All fluorescent stainings were mounted in Fluoromount G (Southern Biotech).

Antibodies were preabsorbed (PA) where noted, and used at the specified final dilutions: rabbit anti-dsRed (1:400; Clontech), rat anti-tropomyosin (1:1,000; Abcam), mouse anti-myosin heavy chain (1:400; gift from S. Abmayr), rabbit anti-Zasp (1:400; gift from F. Schöck), chicken anti-β-gal (1:1,000; Abcam), mouse anti-α-tubulin (1:500; Sigma), rat anti-ensconsin (1:100; gift from P. Rorth), guinea pig anti-Krüppel (PA; 1:2,000; gift from J. Reinitz), rabbit anti-Eve (PA; 1:3,000; gift from M. Frasch), mouse anti-β-PS integrin (1:50; DSHB), rabbit anti-vestigial (PA; 1:50; gift from S. Carroll), fluorescein isothiocyanate-conjugated anti-horseradish peroxidase (1:500; Jackson ImmunoResearch), mouse anti-discs large (1:200; DSHB), rabbit anti-ATP synthase<sup>29</sup> (1:100; gift from H. Duan), rat anti-DE-cadherin (1:100; DSHB), mouse anti-chaoptin 24B10 (1:200; DSHB), rat anti-elav (1:50; DSHB), mouse anti-myosin heavy chain MF20 (1:500; DSHB), rabbit anti-KHC (1:2,500; Santa Cruz), mouse anti-c-Myc (1:5,000; Roche) and rabbit anti-GFP (1:2,500; Invitrogen). Biotinylated secondary antibodies (Vector Laboratories and Jackson ImmunoResearch) and the Vectastain ABC kit (Vector Laboratories) were applied for non-fluorescent stainings. Alexa Fluor 488-conjugated, Alexa Fluor 555-conjugated and Alexa Fluor 647-conjugated secondary antibodies, Alexa Fluor 546-conjugated and Alexa Fluor 647-conjugated phalloidin, Alexa Fluor 488-conjugated wheatgerm agglutinin (Invitrogen), and Hoechst 33342 or 4',6-diamidino-2-phenylindole were used for fluorescent staining (Invitrogen).

Fluorescent images of *Drosophila* were acquired on a Leica SP5 laser scanning confocal microscope running the LAS AF 2.2 software (objectives used: 20× 0.70 numerical aperture (NA) HC PL APO multi-immersion, 40× 1.25 NA, 63× 1.4 NA or 100× 1.43 NA HCX PL APO oil-immersion). Non-fluorescent images were acquired on a Zeiss Axiophot microscope. Image stacks were analysed and processed using Volocity (Improvision), ImageJ, and Photoshop CS4 (Adobe).

Fluorescent images of mouse cells were acquired with a Nikon Ti microscope equipped with a CoolSNAP HQ2 camera (Roper Scientific), an xy motorized stage (Nikon) and a 40× 1.0 NA PL APO oil-immersion objective, using Metamorph Software (Molecular Devices).

**Quantification methods.** Nuclear positioning was quantified in mammalian myotubes containing at least three nuclei, and myotubes were classified as follows: 'aligned', more than 70% of the nuclei in a myotube aligned along the same axis; 'aggregated', more than 70% of the nuclei did not align along the same axis; 'other', nuclei in a myotube did not fall into either category. Fusion index was quantified as described<sup>30</sup>.

A nearest-neighbour analysis in *Drosophila* was conducted to determine the average distance between a nucleus and its single closest neighbour within a muscle. To identify the closest neighbour for each nucleus in a muscle, the distance between the centre of a nucleus and the centre of each of the closely surrounding nuclei (usually four to eight) was measured. The shortest distance identified the closest nucleus or 'nearest neighbour' and this distance was recorded. A nearest neighbour was identified for each nucleus in every muscle analysed and the nearest neighbour distances were averaged to give an average 'nearest-neighbour' distance for the average space between one nucleus and the next closest nucleus in a muscle.

Sarcomere length was defined as the distance between the Z-lines labelled with the anti-ZASP antibody<sup>15</sup>. Sarcomere length was measured at five separate regions per muscle and in the same five regions for each muscle examined.

Neuromuscular junctions were revealed by staining with horseradish peroxidase or Dlg, and quantified as previously described<sup>17</sup>.

The position of the oocyte nucleus was determined by measuring the distance from the anterior border of the oocyte to the centre of the nucleus.

**Time-lapse imaging.** *Drosophila* embryos were prepared as described<sup>4,38</sup>. Time-lapse sequences were acquired using a Zeiss Axio Imager.Z1 with a 20× 0.75 NA Plan-Apochromat dry objective. All time-lapse series were taken as a set of z-stacks over time (four-dimensional imaging) with optical sections every 3 µm. Only a single z-section with the greatest number of nuclei in focus was selected for each time point. Images were processed with ImageJ and compiled into movie by using Apple Quicktime. Single images of live embryos were acquired on a Zeiss AxioPhot microscope.

Time-lapse imaging of mouse cells was done as reported<sup>30</sup>.

**Yeast two-hybrid screen.** Yeast two-hybrid screening was based on a full-length Ens as bait and performed by Hybrigenics S.A Services using a 0–24-h embryonic *Drosophila* cDNA library; 99 million interactions were analysed, and 10% of the high-confidence clones that were recovered from the screen were Khc.

**Larval behaviour.** Larval behaviour was assessed as described previously, with minor modifications<sup>14,39</sup>. In brief, embryos (stages 16 and 17) were selected for the presence of clustered *apRed* nuclei and/or the absence of the fluorescent balancer. Selected embryos were placed on a yeast-coated apple-juice plate overnight at 25 °C, and L1 larvae were selected on the following day and placed into vials of standard food containing bromophenol blue. Larvae were picked from the vial 3 days later and tracked. Larvae were tracked individually as they migrated towards a single odour source (0.25 M ethyl butyrate; Sigma) and recorded for 5 min with a charge-coupled device camera until they reached the odour source or until they contacted any of the walls of the apparatus. Images were processed by Ethovision software (Noldus).

**Germline transformation and constructs.** *UAS-ensHA* DNA was constructed by PCR-amplifying the full-length cDNA from the *ens* RA transcript with primers to add a haemagglutinin tag to the C-terminal end of the protein followed by an *EcoRI* restriction site. Primers used were 5'-GCCGAATTCCACCATGGCGA GTCTTGGGGGC-3' (5') and 5'-GCCGAATTCTTATCAAGCGTAATCTGGA ACATCGTATGGGTACAGCAGCGATATATCTTTATTTTCGTG-3' (3'). Amplified cDNA was introduced into the Uni-ZAP XR vector (a derivative of pBlueScript SK<sup>-</sup>; Stratagene) using *EcoRI/XhoI*. This DNA was used by BestGene Inc. to generate transgenic flies.

**Statistics.** Statistical analysis was performed with Prism (GraphPad Software Inc.). Pairwise comparisons were made with Student's *t*-test, and group comparisons were made by analysis of variance followed by Tukey's posthoc analysis. In nuclear positioning analysis in C2C12 cells, Student's *t*-tests were performed between scrambled siRNA and experimental condition for each category (aligned, aggregated and other).

31. Halfon, M. S. *et al.* New fluorescent protein reporters for use with the *Drosophila* Gal4 expression system and for vital detection of balancer chromosomes. *Genesis* **34**, 135–138 (2002).
32. Lewis, E. *Method of Feeding Ethane Methylsulfonate (EMS) to Drosophila Males*. (Drosophila Information Service, 1968).
33. Parlakian, A. *et al.* Skeletal muscle phenotypically converts and selectively inhibits metastatic cells in mice. *PLoS ONE* **5**, e9299 (2010).
34. De Palma, C. *et al.* Nitric oxide inhibition of Drp1-mediated mitochondrial fission is critical for myogenic differentiation. *Cell Death Differ.* **17**, 1684–1696 (2010).
35. Khatau, S. B. *et al.* A perinuclear actin cap regulates nuclear shape. *Proc. Natl Acad. Sci. USA* **106**, 19017–19022 (2009).
36. Brent, J., Werner, K. & McCabe, B. *Drosophila* larval NMJ dissection. *J. Vis. Exp.* 10.3791/1107 (2009).
37. Brent, J., Werner, K. & McCabe, B. D. *Drosophila* larval NMJ immunohistochemistry. *J. Vis. Exp.* (25) e1108, doi:10.3791/1108 (2009).
38. Richardson, B. E., Beckett, K. & Baylies, M. K. Live imaging of *Drosophila* myoblast fusion. *Methods Mol. Biol.* **475**, 263–274 (2008).
39. Louis, M., Piccinotti, S. & Vosshall, L. B. High-resolution measurement of odor-driven behavior in *Drosophila* larvae. *J. Vis. Exp.* (11) e638, doi:10.3791/638 (2008).

# Small-molecule inhibitors of the AAA+ ATPase motor cytoplasmic dynein

Ari J. Firestone<sup>1\*</sup>, Joshua S. Weinger<sup>2\*</sup>, Maria Maldonado<sup>2</sup>, Kari Barlan<sup>3</sup>, Lance D. Langston<sup>2</sup>, Michael O'Donnell<sup>2</sup>, Vladimir I. Gelfand<sup>3</sup>, Tarun M. Kapoor<sup>2</sup> & James K. Chen<sup>1</sup>

**The conversion of chemical energy into mechanical force by AAA+ (ATPases associated with diverse cellular activities) ATPases is integral to cellular processes, including DNA replication, protein unfolding, cargo transport and membrane fusion<sup>1</sup>. The AAA+ ATPase motor cytoplasmic dynein regulates ciliary trafficking<sup>2</sup>, mitotic spindle formation<sup>3</sup> and organelle transport<sup>4</sup>, and dissecting its precise functions has been challenging because of its rapid timescale of action and the lack of cell-permeable, chemical modulators. Here we describe the discovery of ciliobrevins, the first specific small-molecule antagonists of cytoplasmic dynein. Ciliobrevins perturb protein trafficking within the primary cilium, leading to their malformation and Hedgehog signalling blockade. Ciliobrevins also prevent spindle pole focusing, kinetochore-microtubule attachment, melanosome aggregation and peroxisome motility in cultured cells. We further demonstrate the ability of ciliobrevins to block dynein-dependent microtubule gliding and ATPase activity *in vitro*. Ciliobrevins therefore will be useful reagents for studying cellular processes that require this microtubule motor and may guide the development of additional AAA+ ATPase superfamily inhibitors.**

The AAA+ superfamily of enzymes couples ATP hydrolysis with the generation of mechanical force to regulate diverse aspects of prokaryote and eukaryote biology<sup>1</sup>. ATP-dependent conformational changes can propagate through these molecular machines to complete cellular processes within seconds, and chemical inhibitors that act rapidly and reversibly are much-needed tools for investigating the cellular functions of individual superfamily members. Yet, so far, only one AAA+ ATPase mechanoenzyme has been selectively targeted by a small molecule<sup>5</sup>.

We recently conducted a high-throughput screen for inhibitors of the Hedgehog (Hh) pathway (Fig. 1a)<sup>6</sup>, a key mediator of embryonic development and oncogenesis<sup>7</sup>. Our study was designed to identify compounds that act downstream of Smoothened (Smo), a transmembrane Hh signalling protein, and one of the small molecules, HPI-4 (Fig. 1b; 1), blocked Hh pathway activation in cells lacking the negative regulator Suppressor of Fused (Sufu)<sup>6</sup>. Prolonged treatment of cells with this benzoyl dihydroquinazolinone also reduced the number and size of primary cilia, a microtubule-based extension of the plasma membrane that is required for Hh signalling<sup>8</sup>. Intrigued by these cellular phenotypes, we investigated the biochemical mechanism of HPI-4.

We first synthesized a series of analogues (Fig. 1b; 2–9) and evaluated their effects on Hh signalling and primary cilia formation (Fig. 1c, d and Supplementary Figs 1–3). Chemical derivatives lacking either a 3- or 4-chloro substituent on the benzoyl ring system (2 and 6) or the acyclic ketone (9) were significantly less active in either assay (Fig. 1b, d). The other small molecules segregated into 2,4-dichlorobenzoyl dihydroquinazolinones that inhibit both Hh signalling and primary cilia formation (1, 3–5), which we henceforth name 'ciliobrevins A–D', and monochlorobenzoyl analogues that can block Hh target gene expression without inducing ciliary defects (7 and 8) (Fig. 1b, d).

Hh signalling is primarily mediated by the transcription factors Gli2 and Gli3, which exist in a pathway state-dependent balance of amino (N)-terminal repressors (Gli2/3R), full-length polypeptides (Gli2/3FL) and transcriptional activators (Gli2/3A) (Fig. 1a)<sup>7</sup>. Both repressor and activator formation require the primary cilium<sup>8</sup>, and accordingly ciliobrevins altered the Gli3FL/Gli3R ratio in cells stimulated with the N-terminal domain of Sonic Hedgehog (Shh-N) (Fig. 1e; 30  $\mu$ M doses of each compound). Shh-N-dependent Gli3FL phosphorylation was also reduced by these compounds, perhaps reflecting loss of Gli3A<sup>9</sup>. In contrast, none of the other analogues had a significant effect on the Gli3 processing or phosphorylation state (Fig. 1e).

To understand better the basis of these phenotypes, we took advantage of the temporal control afforded by chemical perturbations. Although prolonged exposure to these compounds causes defects in axonemal morphology, shorter treatments can divulge ciliobrevin-sensitive processes within structurally intact cilia. Because Hh pathway activation coincides with Gli2 accumulation at the distal ciliary tip<sup>10</sup>, we examined the effect of ciliobrevins on Gli2 localization (Fig. 1f). We incubated Hh-responsive cells with individual compounds at a 30  $\mu$ M concentration in the absence or presence of Shh-N-conditioned medium for 4 h. Gli2 localization was unchanged by derivatives that do not significantly perturb ciliogenesis (2 and 8), whereas ciliobrevins A and D (1 and 5) induced ciliary Gli2 levels comparable to that in Shh-N-stimulated cells.

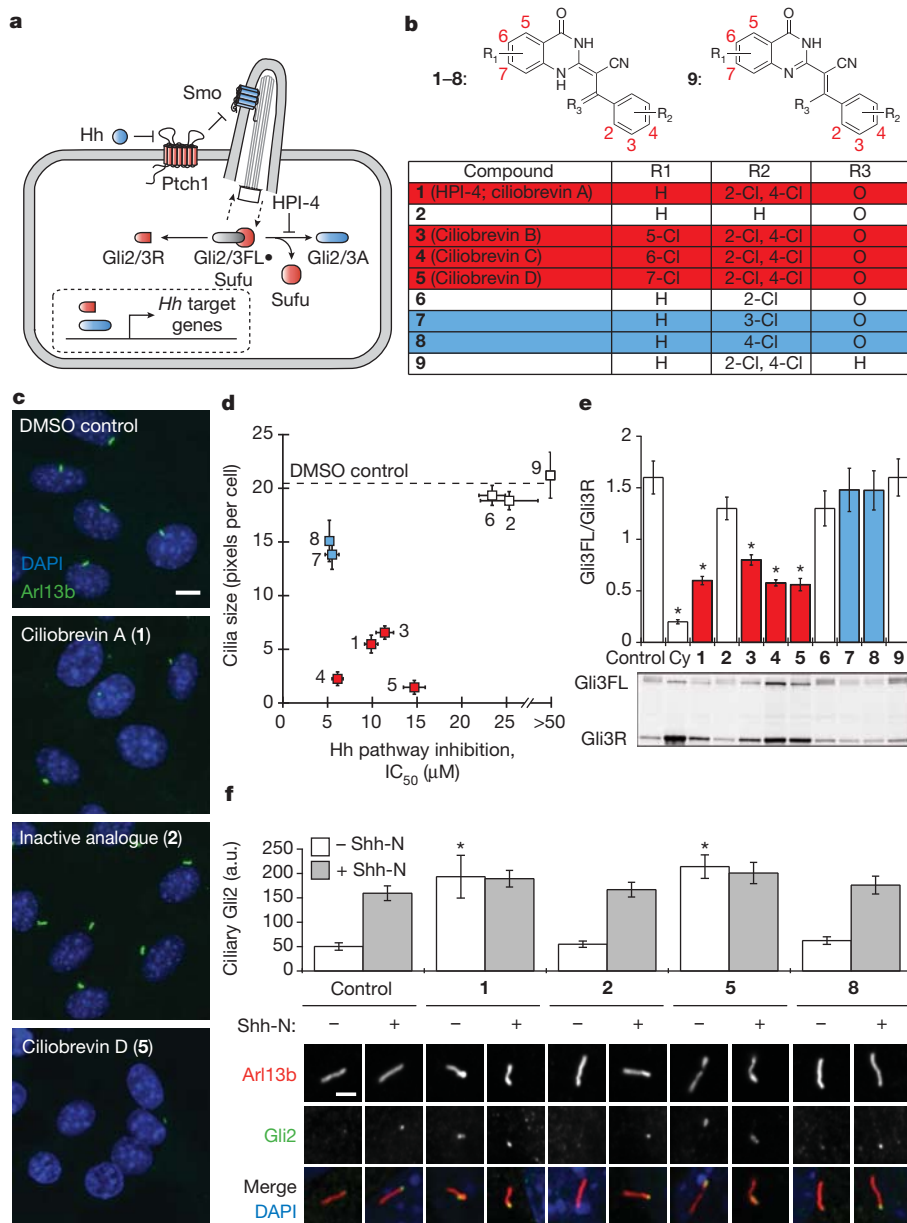
The ability of ciliobrevins to increase ciliary Gli2 levels suggests that these compounds might target protein trafficking mechanisms within this organelle. Intraflagellar transport (IFT) can be resolved into anterograde trafficking, which requires the plus-end-directed motor kinesin-2 and the IFTB multisubunit complex, and retrograde trafficking, which uses the minus-end-directed motor cytoplasmic dynein 2 and the IFTA complex<sup>8</sup>. Loss of the primary cilia-specific cytoplasmic dynein 2 heavy chain (Dync2h1) alters cilia morphology<sup>11</sup>, reduces Hh target gene expression<sup>11</sup> and increases ciliary levels of Gli2<sup>10</sup>. Similarities between these genetic phenotypes and the effects of ciliobrevins led us to hypothesize that these small molecules might inhibit cytoplasmic dynein 2. We therefore examined the effect of ciliobrevins on the subcellular localization of IFTB component IFT88, which requires cytoplasmic dynein 2-dependent retrograde transport for its return to the basal body. Treating cells for 1 hour with ciliobrevin D (5) but not DMSO or an inactive analogue (2) significantly increased IFT88 levels at the distal tip of primary cilia (Supplementary Fig. 4), providing further evidence that ciliobrevins inhibit cytoplasmic dynein 2 function.

Cytoplasmic dynein complexes have other cellular functions, including the crosslinking and focusing of microtubule minus ends within the mitotic spindle<sup>3</sup>. These actions create the fusiform shape and localize  $\gamma$ -tubulin-containing complexes to the spindle poles<sup>3</sup>. Cytoplasmic dynein 1 inhibition by blocking antibodies or dominant-negative constructs perturbs spindle assembly, resulting in disorganized poles and reduced  $\gamma$ -tubulin recruitment<sup>3,12–14</sup>. To determine whether

<sup>1</sup>Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, California 94305, USA. <sup>2</sup>Laboratory of Chemistry and Cell Biology, Rockefeller University, New York City, New York 10021, USA. <sup>3</sup>Department of Cell and Molecular Biology, Northwestern University School of Medicine, Chicago, Illinois 60611, USA.

\*These authors contributed equally to this work.





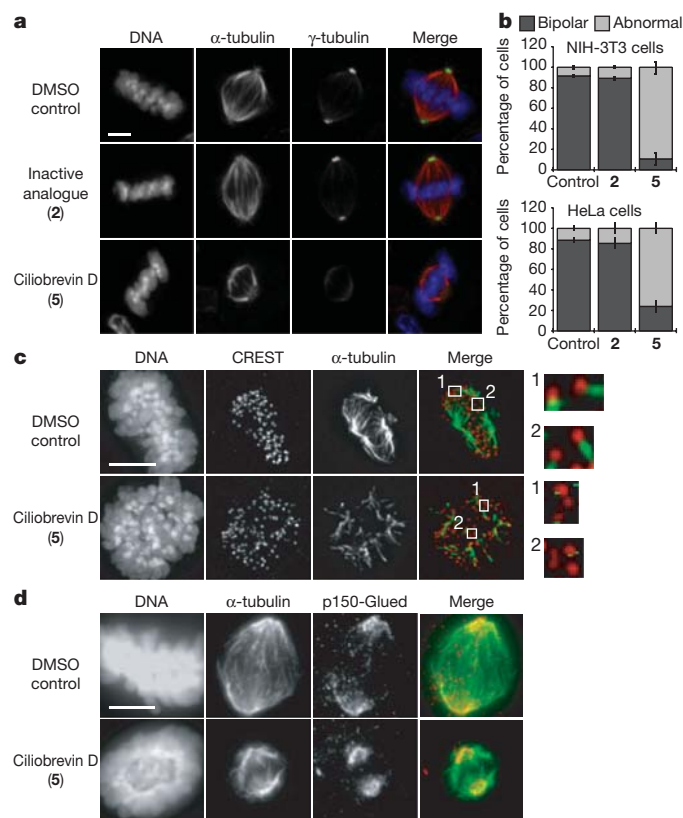
**Figure 1 | Ciliobrevins disrupt primary cilium-dependent Gli regulation.**  
**a**, Depiction of Hh signalling with positive (blue) and negative (red) regulators.  
**b**, Dihydroquinazolinone structures. **c**, Ciliary phenotypes after treatment with individual compounds (30  $\mu$ M) for 24 h. Scale bar, 10  $\mu$ m. **d**, Compound effects on Hh target gene expression in Shh-LIGHT2 cells<sup>30</sup> and cilia size in Shh-EGFP cells<sup>6</sup>. Half-maximum inhibitory concentration (IC<sub>50</sub>) values are the average of three independent experiments  $\pm$  s.d., and cilia size is defined as the average number of Arl13b-positive pixels per nuclei for ten micrographs  $\pm$  s.d., each containing approximately 150 cells. Small molecules that block Hh signalling alone (blue) and those that inhibit both Hh pathway activity and ciliogenesis

(ciliobrevins; red) are highlighted. **e**, Gli3FL and Gli3R levels in Shh-EGFP cells treated with Shh-N and either individual dihydroquinazolinones (30  $\mu$ M), 3  $\mu$ M cyclopamine or DMSO for 16 h. Average Gli3FL/Gli3R ratios for five independent experiments  $\pm$  s.e.m. and a representative immunoblot are shown, with asterisks indicating  $P < 0.005$  for individual compounds versus DMSO and cyclopamine (Cy). **f**, Ciliary Gli2 levels (a.u., arbitrary units) in Shh-EGFP cells treated with selected ciliobrevins (**1** and **5**), inactive analogues (**2** and **8**) or DMSO for 4 h. Average Gli2 levels in the distal end of at least 25 cilia  $\pm$  s.e.m. and representative confocal micrographs are shown. Asterisks indicate  $P < 0.005$  for individual compounds versus DMSO. Scale bar, 2  $\mu$ m.

ciliobrevins recapitulate these phenotypes, we treated a metaphase-enriched population of NIH-3T3 cells with 50  $\mu$ M of either ciliobrevin D (**5**) or an inactive analogue (**2**) for 1 hour and examined their mitotic structures. Cells treated with ciliobrevin D showed abnormal (unfocused, multipolar or collapsed) spindles with disrupted  $\gamma$ -tubulin localization (Fig. 2a, b and Supplementary Fig. 5a), whereas cells incubated with the non-cilia-disrupting analogue or vehicle alone showed normal spindle morphologies. Similar ciliobrevin-induced spindle defects were observed in HeLa cells, although to a lesser extent (Fig. 2b). Cytoplasmic dynein 1 is also required for establishing stable kinetochore-microtubule interactions<sup>15</sup>, and ciliobrevin D treatment disrupted the formation of

cold-stable microtubules that mediate proper spindle-chromosome attachments (Fig. 2c).

To investigate if these spindle-disruptive effects were associated with altered dynein localization, we examined binding partners that recruit or co-localize with this motor. Immunofluorescence microscopy showed that p150-Glued, a dynactin component proposed to recruit dynein to kinetochores and spindle poles during mitosis<sup>16</sup>, was localized to the disorganized spindle poles of ciliobrevin D-treated cells (Fig. 2d and Supplementary Fig. 5b). Kinetochores targeting of p150-Glued was also unaffected by ciliobrevin D, as assessed in nocodazole-treated cells to obviate effects due to microtubule-attachment status



**Figure 2 | Ciliobrevins disrupt spindle pole assembly and kinetochore-microtubule attachment.** **a**, Mitotic spindles observed in NIH-3T3 cells treated with MG132 for 90 min and subsequently cultured with either an inactive analogue (2) or ciliobrevin D (5) at a 50  $\mu$ M dose or DMSO for 1 h. Staining for DNA,  $\alpha$ -tubulin and  $\gamma$ -tubulin is shown. **b**, Quantification of spindle phenotypes in NIH-3T3 and HeLa cells treated as described above and scored for either bipolar or abnormal morphologies. Data are the average of three independent experiments  $\pm$  s.e.m., each including at least 150 spindles. **c**, Kinetochore-microtubule interactions analysed in metaphase-arrested NIH-3T3 cells treated with DMSO or 50  $\mu$ M 5 and then incubated on ice for 10 min. Staining for DNA, the kinetochore marker CREST and  $\alpha$ -tubulin is shown. Insets highlight individual kinetochore-microtubule attachments or untethered kinetochores (400 $\times$  magnification). **d**, Localization of p150-Glued in metaphase-arrested NIH-3T3 cells treated with DMSO or 50  $\mu$ M 5. Staining for DNA,  $\alpha$ -tubulin and p150-Glued is shown. Scale bars: **a**, 4  $\mu$ m; **c**, **d**, 5  $\mu$ m.

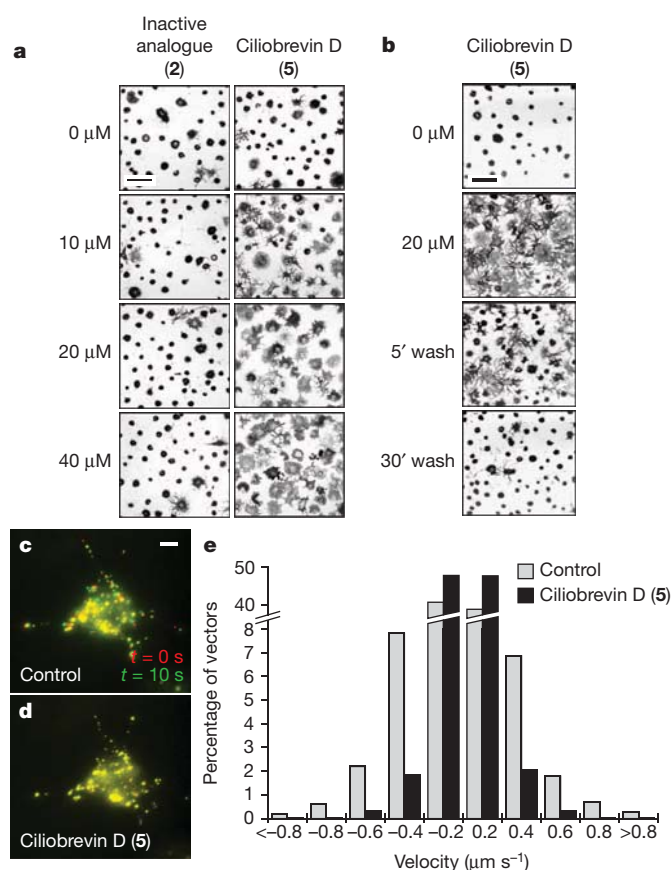
(Supplementary Fig. 6). We similarly observed that Zeste white 10 (Zw10), a component of the Rod/Zw10/Zwlich complex that recruits cytoplasmic dynein to kinetochores<sup>17</sup>, and the kinetochore-associated protein Centromere protein E (CENP-E)<sup>18</sup> correctly localized to kinetochores under these conditions (Supplementary Fig. 6), indicating that their recruitment and the kinetochore structure itself are not disrupted by ciliobrevins. Thus, ciliobrevin-induced spindle phenotypes most likely result from dynein inhibition rather than mislocalization.

To characterize further the mitotic defects associated with ciliobrevin treatment, we conducted real-time confocal microscopy of green fluorescent protein (GFP)-tubulin-expressing NIH-3T3 cells (Supplementary Figs 7, 8). Within minutes of ciliobrevin D treatment, spindles collapsed, bipolarity was lost and spindle poles appeared disorganized (Supplementary Fig. 7b). Upon compound washout, bipolar spindles quickly re-emerged and chromosomes segregated at anaphase without any pronounced defects (Supplementary Fig. 7c, d). Ciliobrevin D addition also reversibly disrupted the pre-formed spindles of metaphase-arrested cells (Supplementary Fig. 8b, c) and reduced overall microtubule levels (Supplementary Fig. 8d). These latter effects are mitosis-specific, as microtubule levels in non-dividing cells were unaffected by inhibitor treatment (Supplementary Fig. 7e, f). Taken together, these results reveal that cytoplasmic dynein is not only

required for spindle pole assembly but also actively participates in its maintenance.

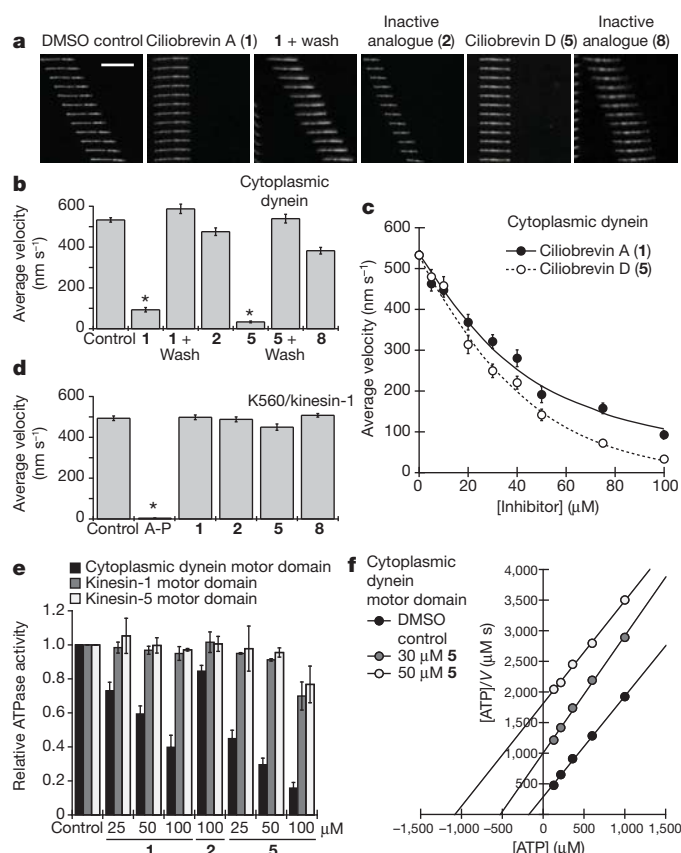
Cytoplasmic dynein 1 also regulates organelle trafficking, such as the melatonin-induced aggregation of melanosomes in *Xenopus* melanophores<sup>19</sup>. To determine if melanosome trafficking is sensitive to ciliobrevins, we cultured *Xenopus* melanophores with melanocyte-stimulating hormone to disperse these pigment granules and then treated the cells with melatonin and various concentrations of ciliobrevin D (5) or an inactive analogue (2). Ciliobrevin D reversibly inhibited melanosome aggregation, but the non-cilia-disrupting derivative had no discernible effect at comparable doses (Fig. 3a, b and Supplementary Movies 1–3). Ciliobrevin D similarly abrogated the movement of peroxisomes in *Drosophila* S2 cells (Fig. 3c–e and Supplementary Movies 4 and 5), consistent with the role of cytoplasmic dynein in their bidirectional motility<sup>20</sup>.

Collectively, these results indicate that ciliobrevins are specific, reversible inhibitors of disparate cytoplasmic dynein-dependent processes. Ciliobrevins do not perturb cellular mechanisms that are independent of dynein function, including actin cytoskeleton organization (Supplementary Fig. 9) and the mitogen-activated protein kinase and phosphoinositol-3-kinase signalling pathways<sup>6</sup>. To examine more directly whether cytoplasmic dynein is the direct target of ciliobrevins, we evaluated their effects on dynein-dependent microtubule gliding *in vitro*. Ciliobrevins A and D (1 and 5) retarded the ATP-dependent



**Figure 3 | Ciliobrevins inhibit melanosome aggregation and peroxisome motility.** **a**, Brightfield images of *Xenopus* melanophores treated with various concentrations of ciliobrevin D (5) or an inactive derivative (2) for 10 min, stimulated with melatonin in the presence of the compounds for 30 min, and then paraformaldehyde-fixed. **b**, Melanophores treated with 5 as before, washed in medium containing melatonin alone, and then imaged. Scale bars, 100  $\mu$ m. **c**, **d**, Motility of GFP-labelled peroxisomes in *Drosophila* S2 cells cultured in the absence (**c**) or presence (**d**) of 5. Overlays of videomicroscopy frames at  $t = 0$  s (red) and  $t = 10$  s (green) are shown. Scale bar, 5  $\mu$ m. **e**, Vector distributions for the GFP-labelled peroxisomes with movement towards and away from the cell centre denoted by negative and positive bin values, respectively.

movement of fluorescently labelled microtubules on bovine cytoplasmic dynein-coated glass slides in a reversible and dose-dependent manner (Fig. 4a–c, Supplementary Fig. 10 and Supplementary Movies 6–8); analogues that did not perturb cytoplasmic dynein-dependent processes in our cell-based assays (2 and 8) had minimal effects (Fig. 4a, b and Supplementary Movie 9). The conserved structure–activity relationships of dihydroquinazolinones in the microtubule gliding and cell-based assays confirm cytoplasmic dynein as the ciliobrevin target, and neither ciliobrevin A nor D significantly affected K560/kinesin-1-dependent microtubule gliding *in vitro* at 100  $\mu\text{M}$  concentrations (Fig. 4d and Supplementary Movies 10–12). The compounds do not broadly target members of the AAA+ ATPase family either, as they have no effect on p97-dependent degradation of endoplasmic-reticulum-associated proteins (Supplementary Fig. 11) or Mcm2–7-mediated DNA unwinding (Supplementary Fig. 12).



**Figure 4 | Ciliobrevins inhibit cytoplasmic dynein-dependent microtubule gliding and ATPase activity.** **a**, Montages of fluorescent microtubules moving on bovine dynein-coated glass slides in the presence of ATP and either DMSO, ciliobrevin A (1), ciliobrevin D (5) or non-cilia-disrupting analogues (2 and 8). All compounds were tested at a 100  $\mu\text{M}$  concentration and the vertically stacked images were acquired 2 s apart. Scale bar, 10  $\mu\text{m}$ . **b**, Quantification of the compounds' effects on dynein-dependent microtubule gliding. Data are the average velocities for at least 56 microtubules  $\pm$  s.e.m. Asterisks indicate  $P < 10^{-6}$  and at least 30% inhibition for individual compounds versus DMSO. **c**, Dose responses of 1 and 5 in the dynein-dependent assay. Data are the average velocities for at least 45 microtubules  $\pm$  s.e.m. **d**, Effects of 1, 2, 5, 8 and the competitive ATP antagonist adenyllyl imidodiphosphate (A-P) on microtubule gliding driven by the K560 N-terminal fragment of kinesin-1. Dihydroquinazolinones and AMP-PNP were tested at a 100  $\mu\text{M}$  and 1 mM concentrations, respectively. Data are the average velocities for at least 34 microtubules  $\pm$  s.e.m., and asterisks indicate  $P < 0.01$  for individual compounds versus DMSO. **e**, Effects of 1, 2 and 5 on the ATPase activities of recombinant motor domains derived from rat cytoplasmic dynein, human kinesin-1 and human kinesin-5. Compound concentrations in micromolar units are shown, and data are the average ATPase activities for two independent experiments  $\pm$  s.d. **f**, Hanes–Woolf plot of rat dynein motor ATPase velocity (V) suggesting the nucleotide-competitive activity of 5.

We next investigated how ciliobrevins abrogate cytoplasmic dynein function. Neither ciliobrevin A nor D was able to disrupt the association between ADP-bound dynein and microtubules, as determined in a co-sedimentation assay (Supplementary Fig. 13). Both small molecules, however, were able to inhibit the ATPase activity of bovine brain cytoplasmic dynein in a concentration-dependent manner, whereas their inactive analogue (2) could not (Supplementary Fig. 14). The compounds had analogous effects on the ATPase activity of recombinant rat dynein motor domain, but none significantly inhibited the ATPase activities of recombinant motor domains derived from human kinesin-1 or kinesin-5 (Fig. 4e). Ciliobrevin efficacies at various ATP concentrations suggest that these small molecules act in a nucleotide-competitive manner (Fig. 4f). Consistent with this mechanism, ciliobrevin A also inhibited ultraviolet-light-induced cleavage of the cytoplasmic dynein motor domain in the presence of sodium vanadate and ATP (Supplementary Fig. 15).

Our studies establish ciliobrevins as the first small molecules known specifically to inhibit cytoplasmic dynein *in vitro* and in live cells. Although the ATP analogue erythro-9-[3-(2-hydroxynonyl)]adenine<sup>21</sup> and the antioxidant nordihydroguaiaretic acid<sup>22</sup> have been previously reported to abrogate dynein function, these compounds are promiscuous enzyme antagonists<sup>23,24</sup>. The natural product purealin can partly inhibit dynein ATPase activity *in vitro*<sup>25</sup>, but its ability to block cytoplasmic dynein-dependent cellular processes has not been demonstrated. Our studies indicate that ciliobrevins can inhibit both cytoplasmic dynein 1 and 2, and, accordingly, the compounds will be broadly applicable probes of dynein-dependent processes. Further development of ciliobrevin-like molecules could lead to isoform-selective inhibitors of this minus-end-directed microtubule motor and perhaps specific antagonists of other AAA+ ATPase superfamily members.

## METHODS SUMMARY

**Hh signalling assays.** Hh signalling and Gli3 processing assays were performed as described<sup>6,9</sup>.

**Cell imaging.** Cilia were immunostained with anti-Arl13b antibody (T. Caspar), and cilia size was determined by dividing the number of Arl13b-positive pixels by the number of nuclei per image. Ciliary levels of Gli2 and IFT88 were determined by immunostaining cells with anti-Gli2 (R&D Systems) or anti-IFT88 (ProteinTech Group) antibodies. Analyses of mitotic spindles and kinetochore–microtubule attachments in fixed cells were performed as described<sup>26</sup>, and real-time confocal microscopy was conducted with NIH-3T3 cells stably expressing GFP-tubulin. Melanosome and peroxisome motility studies were conducted as reported<sup>19,20</sup>.

**Dynein activity assays.** Bovine brain dynein was purified as described<sup>27</sup>, and its microtubule gliding activity was assayed as reported<sup>28</sup> with modifications. Recombinant rat cytoplasmic dynein motor domain was heterologously expressed and purified as described<sup>29</sup>. ATPase activities were determined using a Malachite Green assay (Novus Biologicals).

Full Methods are in the Supplementary Information.

Received 15 June 2011; accepted 1 February 2012.

Published online 18 March 2012.

- White, S. R. & Lauring, B. AAA+ ATPases: achieving diversity of function with conserved machinery. *Traffic* **8**, 1657–1667 (2007).
- Scholey, J. M. Intraflagellar transport. *Annu. Rev. Cell Dev. Biol.* **19**, 423–443 (2003).
- Merdes, A., Ramyar, K., Vechio, J. D. & Cleveland, D. W. A complex of NuMA and cytoplasmic dynein is essential for mitotic spindle assembly. *Cell* **87**, 447–458 (1996).
- Akhmanova, A. & Hammer, J. A. III. Linking molecular motors to membrane cargo. *Curr. Opin. Cell Biol.* **22**, 479–487 (2010).
- Chou, T. F. et al. Reversible inhibitor of p97, DBE9, impairs both ubiquitin-dependent and autophagic protein clearance pathways. *Proc. Natl Acad. Sci. USA* **108**, 4834–4839 (2011).
- Hyman, J. M. et al. Small-molecule inhibitors reveal multiple strategies for Hedgehog pathway blockade. *Proc. Natl Acad. Sci. USA* **106**, 14132–14137 (2009).
- Jiang, J. & Hui, C.-C. Hedgehog signaling in development and cancer. *Dev. Cell* **15**, 801–812 (2008).
- Goetz, S. C. & Anderson, K. V. The primary cilium: a signalling centre during vertebrate development. *Nature Rev. Genet.* **11**, 331–344 (2010).



9. Humke, E. W., Dorn, K. V., Milenkovic, L., Scott, M. P. & Rohatgi, R. The output of Hedgehog signaling is controlled by the dynamic association between Suppressor of Fused and the Gli proteins. *Genes Dev.* **24**, 670–682 (2010).
10. Kim, J., Kato, M. & Beachy, P. A. Gli2 trafficking links Hedgehog-dependent activation of Smoothened in the primary cilium to transcriptional activation in the nucleus. *Proc. Natl Acad. Sci. USA* **106**, 21666–21671 (2009).
11. Huangfu, D. & Anderson, K. V. Cilia and Hedgehog responsiveness in the mouse. *Proc. Natl Acad. Sci. USA* **102**, 11325–11330 (2005).
12. Heald, R. *et al.* Self-organization of microtubules into bipolar spindles around artificial chromosomes in *Xenopus* egg extracts. *Nature* **382**, 420–425 (1996).
13. Gaglio, T., Dionne, M. A. & Compton, D. A. Mitotic spindle poles are organized by structural and motor proteins in addition to centrosomes. *J. Cell Biol.* **138**, 1055–1066 (1997).
14. Young, A., Dictenberg, J. B., Purohit, A., Tuft, R. & Doxsey, S. J. Cytoplasmic dynein-mediated assembly of pericentrin and gamma tubulin onto centrosomes. *Mol. Biol. Cell* **11**, 2047–2056 (2000).
15. Varma, D., Monzo, P., Stehman, S. A. & Vallee, R. B. Direct role of dynein motor in stable kinetochore-microtubule attachment, orientation, and alignment. *J. Cell Biol.* **182**, 1045–1054 (2008).
16. King, S. J., Brown, C. L., Maier, K. C., Quintyne, N. J. & Schroer, T. A. Analysis of the dynein-dynactin interaction *in vitro* and *in vivo*. *Mol. Biol. Cell* **14**, 5089–5097 (2003).
17. Starr, D. A., Williams, B. C., Hays, T. S. & Goldberg, M. L. ZW10 helps recruit dynactin and dynein to the kinetochore. *J. Cell Biol.* **142**, 763–774 (1998).
18. Yen, T. J., Li, G., Schaar, B. T., Szilak, I. & Cleveland, D. W. CENP-E is a putative kinetochore motor that accumulates just before mitosis. *Nature* **359**, 536–539 (1992).
19. Gross, S. P. *et al.* Interactions and regulation of molecular motors in *Xenopus* melanophores. *J. Cell Biol.* **156**, 855–865 (2002).
20. Kim, H. *et al.* Microtubule binding by dynactin is required for microtubule organization but not cargo transport. *J. Cell Biol.* **176**, 641–651 (2007).
21. Bouchard, P., Penningroth, S. M., Cheung, A., Gagnon, C. & Bardin, C. W. erythro-9-[3-(2-Hydroxynonyl)]adenine is an inhibitor of sperm motility that blocks dynein ATPase and protein carboxylmethylase activities. *Proc. Natl Acad. Sci. USA* **78**, 1033–1036 (1981).
22. Arasaki, K., Tani, K., Yoshimori, T., Stephens, D. J. & Tagaya, M. Nordihydroguaiaretic acid affects multiple dynein-dynactin functions in interphase and mitotic cells. *Mol. Pharmacol.* **71**, 454–460 (2007).
23. Schliwa, M., Ezzell, R. M. & Euteneuer, U. Erythro-9-[3-(2-hydroxynonyl)]adenine is an effective inhibitor of cell motility and actin assembly. *Proc. Natl Acad. Sci. USA* **81**, 6044–6048 (1984).
24. Park, S., Lee, D. K. & Yang, C. H. Inhibition of fos-jun-DNA complex formation by dihydroguaiaretic acid and *in vitro* cytotoxic effects on cancer cells. *Cancer Lett.* **127**, 23–28 (1998).
25. Zhu, G. *et al.* Synthesis and biological evaluation of purealin and analogues as cytoplasmic dynein heavy chain inhibitors. *J. Med. Chem.* **49**, 2063–2076 (2006).
26. Maldonado, M. & Kapoor, T. M. Constitutive Mad1 targeting to kinetochores uncouples checkpoint signalling from chromosome biorientation. *Nature Cell Biol.* **13**, 475–482 (2011).
27. Woehlke, G. *et al.* Microtubule interaction site of the kinesin motor. *Cell* **90**, 207–216 (1997).
28. Kapoor, T. M. & Mitchison, T. J. Allele-specific activators and inhibitors for kinesin. *Proc. Natl Acad. Sci. USA* **96**, 9106–9111 (1999).
29. Hook, P. *et al.* Long range allosteric control of cytoplasmic dynein ATPase activity by the stalk and C-terminal domains. *J. Biol. Chem.* **280**, 33045–33054 (2005).
30. Taipale, J. *et al.* Effects of oncogenic mutations in Smoothened and Patched can be reversed by cyclopamine. *Nature* **406**, 1005–1009 (2000).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank T. Caspary for anti-Arl13b antibodies, W. Brinkley for human CREST anti-serum, T. Yen for anti-CENP-E antibodies, U. Peters for purified bovine dynein, S. Wacker for human kinesin-5 motor domain, R. Vallee for a pVL1393 baculovirus expression vector containing the rat dynein motor domain, and K. Bersuker and R. Kopito for TCR- $\alpha$ -GFP-expressing cells. This work was supported by funding from the National Institutes of Health (R01 CA136574 to J.K.C.; R01 GM65933 to T.M.K.; R01 GM71772 to T.M.K. and V.I.G.; R01 GM52111 to V.I.G.).

**Author Contributions** J.K.C. and T.M.K. conceived and directed the study. A.J.F. performed chemical syntheses and assays of Hedgehog signaling, primary cilia formation and function, ATPase activity, vanadate-dependent dynein photocleavage and p97-dependent protein degradation. A.J.F. and M.M. performed mitotic spindle analyses. K.B. and V.I.G. designed and interpreted the melanophore and peroxisome trafficking assays. J.S.W. performed microtubule gliding and dynein/microtubule binding assays. L.D.L. and M.O. designed and interpreted the Mcm2–7 helicase assays. A.J.F. and J.K.C. wrote the manuscript with contributions from all other authors.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to J.K.C. ([jameschen@stanford.edu](mailto:jameschen@stanford.edu)) or T.M.K. ([kapoor@rockefeller.edu](mailto:kapoor@rockefeller.edu)).

# CAREERS

**TURNING POINT** Taxonomist's love of insects leads her to genomics research **p.133**

**NATURE NETWORK** Catch up on the latest science-career issues [www.network.nature.com](http://www.network.nature.com)

**NATUREJOBS** For the latest career listings and advice [www.naturejobs.com](http://www.naturejobs.com)



IMAGEZOO/CORBIS

## ETHICS

# The power of transparency

*Public and private institutions have added new rules to ensure transparency and reveal conflicts of interest. For many, following the rules has become harder.*

BY SARAH KELLOGG

Navigating the labyrinth of ever-changing conflict-of-interest rules in science research is resulting in increasing paperwork and lost time for scientists and research institutions. Many well-intentioned individuals are trying to balance the demands of research enterprise with maintaining public

confidence and scientific objectivity.

The heightened sensitivity is a result, in part, of a series of scandals and the growing number of links between academic researchers and industry. Governments, academic institutions and research centres are fine-tuning their financial conflict-of-interest policies to strengthen disclosure requirements and place limits on industry payments, all in the interest

of promoting transparency.

"It's a whole new game in terms of financial relationships," says David Rothman, director of the Center for the Study of Science and Medicine at Columbia University in New York. "It's unprecedented. If the information can be found, your dean is going to know. The journal editor is going to know. The National Institutes of Health is going to know. And if everyone knows, it means you had better be pretty thorough in disclosing and managing your conflicts."

Exhaustive reporting is the best strategy to maintain reputations and to protect researchers from charges of bias, say conflict-of-interest experts. Senior faculty members, early-career researchers and medical students should make sure that they leave little room for doubt about their financial relationships with industry.

Yet this zeal for disclosure poses a significant challenge for both institutions and individuals. Complying with the rules is expected to cost institutions millions of dollars, and is a considerable intrusion into scientists' time and privacy. Comprehensive reporting can leave the impression among members of the public that a 'conflict' exists when, in fact, the authors only have a competing or allied interest.

## NEW RULES, NEW RESPONSIBILITIES

The international crackdown on disclosing all industry payments comes in the wake of a series of controversial cases. These include a 2008 US congressional investigation into researchers from Harvard University in Cambridge, Massachusetts, who failed to report millions in consulting fees from drug manufacturers; a June 2011 report from the news agency Reuters that revealed how scientists whose research was published in the *British Journal of Dermatology* failed to disclose even the most blatant financial conflicts on the publication forms; and a 2011 decision by the European Parliament to audit the European Medicines Agency (EMA) after its former executive director, Thomas Lönngren, accepted a paid consultancy with a market-access company soon after leaving his position.

The US Department of Health and Human Services in Washington DC staked out new territory last August, releasing revamped financial conflict-of-interest rules, which updated disclosure regulations that have been in place since 1995. The rules take effect in August 2012 and will affect more than 40,000 researchers.

Under the regulations, the threshold for disclosing industry and non-profit payments or gifts to researchers and their immediate family members decreases from US\$10,000 ►

► to \$5,000. Violations will be punishable by suspension or by the termination of federal research funding. Academic and research institutions will be required to strengthen their disclosure activities by developing more robust regulations, as well as by expanding their systems to educate staff and report conflicts to regulators and the public.

"There are, increasingly, more interactions with industry as the knowledge base grows and there is more translation of that knowledge into practical outcomes," says Sally Rockey, deputy director for extramural research at the US National Institutes of Health (NIH) in Bethesda, Maryland. "We wanted to take a look at our regulations to make sure they supported investigators, as appropriate, as well as our ability to oversee the process."

In September 2011, the EMA released rules requiring scientific experts advising on drug approvals to provide a signed declaration detailing their direct and indirect financial interests, as well as any other interests that could influence them. The EMA also created a searchable database to allow the public to track those declarations. Some scientists expressed concern that the policy might have an effect on the number of experts able to serve on government advisory boards because of the sheer number of

relationships between investigators and industry. Newer fields, such as nanotechnology or genetic engineering, are particularly vulnerable.

Moves to strengthen the reporting of conflicts of interest come at a crucial time financially for institutions such as academic research centres. With budgets forcing reductions in government spending on science research around

**"We try to promote the idea that having a conflict of interest is not necessarily a negative."**

the world, industry is increasingly stepping in to fill the gap, especially in translational and clinical studies. Although scientists would be wise to proceed with caution, these partnerships can be fruitful. "We don't want to inhibit these relationships," says Rockey. "We know they're there." Researchers should not assume that receiving industry money will automatically block an appointment to an NIH study group, a drug-approval committee or a formulary committee at a medical centre, Rockey adds (see 'Staying clean').

Clearly, the rules of the game differ considerably depending on the institution or country. Ethics experts therefore recommend that the first place to visit in the disclosure process

should be the office for ethics and compliance at the researcher's institution. Most academic institutions have offices and committees charged with managing these issues, and everyone from experienced investigators to graduate students have a responsibility to know the rules. Reporting every conflict of interest may be intrusive, and some researchers may worry that it suggests an unproven tie between industry fees and research findings; but disclosure is the first and best defence against bias charges, and not disclosing information could look suspicious.

Some institutions not only offer one-on-one counselling but also provide helpful seminars and online tutorials to explain the rules and disclosure processes. "We offer many opportunities for our investigators to learn more about their responsibilities in conflict-of-interest disclosure and to ensure they aren't letting bias enter into investigations," says Guy Chisolm, director of the Innovation Management and Conflict of Interest Program at the Cleveland Clinic, an academic medical-research centre in Ohio. "This kind of transparency is a baseline for a change of culture and a change in attitude."

Armed with government and institutional regulations, researchers need to assess their potential conflicts of interest thoroughly. Have they taken any payments, gifts, speaking fees, meals or educational travel from a private company? What were the amounts and when? They should consider hard questions before accepting an industry-board appointment or attending industry-sponsored events: is this necessary to further their research or career goals?

While institutions determine if a conflict exists, researchers may wrestle with whether to report certain financial contacts, such as direct payments to investigators for the enrolment of participants; income received for work that is tangentially related after the research was completed; and income that is not directly related to the research, such as being on a panel (see 'Daring to disclose'). Almost everyone agrees that it is a major conflict to allow industry to ghostwrite articles, reports or presentations — still a common practice. Investigators who have patents and who conduct classified research or work with for-profit companies or non-profit organizations that have a stated agenda should disclose those financial relationships in detail. Junior researchers should be alert to any potential financial conflicts of their principal investigator, which could have an effect on the young researcher's standing if problems arise.

Even with full disclosure, many doubt the feasibility of removing all bias. Instead of overt reward, what remains could be an unspoken bias that is just as pernicious. "Consciously or not, money can bring with it a need for reciprocity," says Rothman. "Drug and device companies are very clever. They are experts at marketing and making connections. If they didn't feel it brought them some advantage, they wouldn't provide it. They're not in the charity business."

## STAYING CLEAN

### *Is abstinence the answer?*

Abstinence from financial relationships with industry is an increasingly popular strategy to avoid the invasive conflict-disclosure process. Early-career researchers worry that the chance to advance their careers might be thwarted by a financial connection to a private company, or the perception that industry ties may have influenced their own research.

"We do have investigators who have decided not to accept any industry money," says Guy Chisolm, director of the Innovation Management and Conflict of Interest Program at the Cleveland Clinic, an academic medical-research centre in Ohio. Chisolm has witnessed a mix of approaches to dealing with conflicts.

But he points out that researchers who take this route choose it themselves, not because the clinic requires them to. The clinic thinks that industry has an important role in speeding up the process from research breakthrough to treatment for patients, and that investigators are capable of having their research funded by industry but still maintaining their objectivity.

Yet many researchers are refusing gifts, travel, education and research grants. "It's easier for them to stay totally clean, so they can get to do whatever they want to do,"

says David Rothman, director of the Center for the Study of Science and Medicine at Columbia University in New York. "The US\$500 or \$1,000 speaking fee just isn't worth it anymore."

The American Medical Student Association in Sterling, Virginia, offers its members the opportunity to sign a PharmFree Pledge and to not accept any payments from the drug industry indefinitely. Although the pledge can send an important signal to employers and funders, it may also severely limit clinical-research opportunities for medical scientists. Some researchers are even avoiding conflict-of-interest situations by eschewing translational research fields in favour of basic-science research.

Many do not endorse such 'zero tolerance' policies. "We really need a call to action to this next generation of scientists," says Michael Kalichman, co-founder and co-director of the Center for Ethics in Science and Technology at the University of California, San Diego. "There is tremendous value to be gained for science in having financial connections with industry. But the nature of those connections is difficult. How do we solve that problem? That is our challenge as a society going forward." **S.K.**



## DARING TO DISCLOSE

## Reporting guidelines

Researchers need to know the rules, to disclose conflicts, and avoid those that are worrisome in the first place.

- Contact the office in your institution or academic centre that handles financial disclosures to determine if you need to make a declaration and what you need to file.
- Familiarize yourself with the reporting rules and disclosure limits of your institution and government.
- Assess your potential financial conflict of interest, including any gifts, speaking fees, meals and educational travel you have accepted.
- Attend seminars or view webinars offered by your institution or government to learn more about the reporting process.
- Be aware of potential industry ties that your principal investigator may have and whether they may reflect badly on you.
- Ask yourself questions about how bias could be introduced into your research and guard against those threats. For example, if you receive industry funding, you should declare this, no matter what the result is.
- Decline industry offers to ghostwrite papers, articles or presentations.
- Carefully assess the advantages and disadvantages of attending industry-sponsored educational events, both as an attendee and as a speaker.
- Consider whether serving on an industry board poses a conflict before accepting.
- Make sure you disclose any financial relationships you may have in all reports, presentations and speeches to ensure transparency.
- If your conflicts of interest change, inform your academic institution in case you need to update your disclosure forms. **S.K.**

Long-term relationships between industry and investigators exist, and will continue to do so. Institutions need a robust strategy to manage these conflicts because no one wants to lose industry investment in research. “This is really about managing conflicts,” says Rockey. “We try to promote the idea that having a conflict of interest is not something bad or necessarily a negative.” ■

**Sarah Kellogg** is a freelance writer in Washington DC.

# TURNING POINT

## Jessica Ware

*Taxonomist and evolutionary biologist Jessica Ware at Rutgers University in Newark, New Jersey, has joined 50 scientists who are using insect genomics to trace the origins of social behaviour, flight and herbivory.*

### Taxonomy is often considered one of the ‘greying’ sciences. What drew you to it?

I have always been interested in the extreme variety in the behaviour and appearance of insects. As I began my graduate work in entomology, I realized that making sense of this diversity was something that had always been done by taxonomists, so I knew that I needed to learn taxonomy as well as systematics, the study of diversity and its origins. I was also aware of the ‘taxonomic impediment’ — there are not enough taxonomists to describe species before they go extinct. So I decided to combine my passion for insects with evolutionary biology.

### How is the field changing?

Taxonomists are moving away from being hyperspecialists who stick to just one taxonomic group — in part because getting funding now requires research to have a broader applicability. So most budding taxonomists are researching a variety of taxonomic groups or exploring broader evolutionary questions. The increased access to genomic sequencing is drawing younger students by offering new ways to combine evolutionary biology with systematics studies that allow us to address broad questions.

### What was your career’s biggest turning point?

I received a postdoctoral research fellowship grant from the US National Science Foundation (NSF), which meant I could focus on what interested me. These grants essentially allow postdocs to pursue their own project, whereas other postdoc positions typically involve working on someone else’s idea. The grant allowed me to hit the ground running and pursue my interests, including molecular dating, which examines how morphology and genetics are related.

### How did you use the grant?

I went to the American Museum of Natural History in New York to work on the systematics of fossil termites, and to the Smithsonian Institution in Washington DC to work on extant species. Being in both places meant I was able to work on fossils and living taxa so I could, for example, paint a realistic picture of how social behaviour evolved in termites. The best part for me was that I was one of very few women and the only African American working as a curator at these museums, so I felt like a trailblazer.



### How have collaborations helped your career?

As a graduate student, I took any project I could that involved collaboration. I worked in Germany, Sweden and South Africa. Every collaboration allowed me to publish outside my PhD research topic. My current collaboration is the result of working with Xin Zhou, who is now at the Beijing Genome Institute (BGI) in China. We were laboratory mates while I was doing my PhD at Rutgers University, and we kept in touch. Among the projects he launched when he moved to the BGI was a large collaborative research initiative looking at insect genomes. He asked me to join. Not all collaborations are easy, but this one seems effortless because everyone is on the same page.

### What does the research project involve?

The 1K Insect Transcriptome Evolution project, or 1KITE, aims to generate transcriptomes — sequences of the expressed genes — of 1,000 species of insect to answer the ‘who’, ‘what’ and ‘why’ questions about insect behaviour. We have begun to collect the species, and eventually we will ask questions about the origins of flight behaviour and egg-laying.

### How will this collaboration help your prospects as a tenure-track faculty member?

One of the biggest pressures for a new assistant professor is getting grants. The success rate for NSF grants, where I go for funding, is low. 1KITE will give me the preliminary genomics data to use as a springboard for future NSF grants, and we expect it to generate good-quality publications in high-impact journals — and that always helps in gaining a tenure package. ■

INTERVIEW BY VIRGINIA GEWIN